(19)日本国特許庁（ＪＰ）　(12)**公開特許公報**（Ａ）　(11)特許出願公開番号
特開2003－150418
（P2003－150418A）
(43)公開日　平成15年５月23日(2003.5.23)

(51)Int.Cl.⁷		識別記号	FI		テーマコード（参考）
Ｇ０６Ｆ	12/00	５１４	Ｇ０６Ｆ	12/00	５１４Ｍ　５Ｂ００５
		５１３			５１３Ｄ　５Ｂ０６５
	3/06	３０２		3/06	３０２Ｊ　５Ｂ０８２
	12/08	５０５		12/08	５０５Ｚ

審査請求　未請求　請求項の数18　ＯＬ　（全 37 頁）

(21)出願番号　特願2001－345525(P2001－345525)

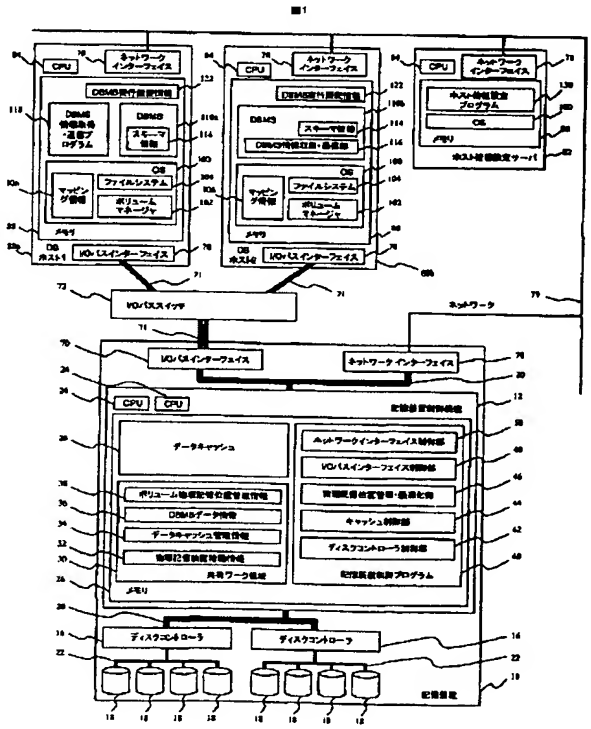
(22)出願日　平成13年11月12日(2001.11.12)

(71)出願人　000005108
株式会社日立製作所
東京都千代田区神田駿河台四丁目６番地
(72)発明者　茂木　和彦
神奈川県川崎市麻生区王禅寺1099番地　株式会社日立製作所システム開発研究所内
(72)発明者　大枝　高
神奈川県川崎市麻生区王禅寺1099番地　株式会社日立製作所システム開発研究所内
(74)代理人　100075096
弁理士　作田　康夫

最終頁に続く

(54)【発明の名称】　データベース管理システムの静的な情報を取得する手段を有する記憶装置

(57)【要約】
【課題】データベース管理システム（ＤＢＭＳ）の特性を考慮したデータ配置やキャッシュの制御を行うことにより、記憶装置のデータアクセス性能を向上させる。
【解決手段】記憶装置は、ネットワーク７９を通してＤＢＭＳの静的な構成情報をＤＢＭＳ情報取得・通信プログラム、ＤＢＭＳ情報通信部、ホスト情報設定プログラムを通して取得し、ＤＢＭＳデータ情報としてメモリ内に記憶する。記憶装置制御プログラム内の物理記憶位置管理・最適化実行部はＤＢＭＳデータ情報を利用してデータ再配置を実行し、キャッシュ制御部はＤＢＭＳデータ情報を加味したデータキャッシュ制御を行う。



【特許請求の範囲】

【請求項１】データベース管理システムが稼動している計算機との接続手段を有し、
前記データベース管理システムにおけるスキーマにより定義される表・索引・ログを含むデータ構造に関する情報と、前記データベース管理システムが管理するデータベースデータを前記スキーマにより定義されるデータ構造毎に分類した前記記憶装置における記録位置に関する情報を取得する情報取得手段を有することを特徴とする記憶装置。

【請求項２】前記接続手段を用いて複数の前記データベース管理システムが稼動している計算機と接続することを特徴とする請求項１に記載の記憶装置。

【請求項３】前記情報取得手段が複数の前記データベース管理システムが管理するデータベースに関する情報を取得することを特徴とする請求項１に記載の記憶装置。

【請求項４】前記情報取得手段は前記接続手段を用いて情報を取得することを特徴とする請求項１に記載の記憶装置。

【請求項５】前記情報取得手段が、前記データベース管理システムが管理するデータベースに関する情報を前記データベース管理システムから取得することを特徴とする請求項１に記載の記憶装置。

【請求項６】前記情報取得手段が、前記データベース管理システムが管理するデータベースに関する情報を前記データベース管理システムとは異なる少なくても１つのプログラムを通して取得することを特徴とする請求項１に記載の記憶装置。

【請求項７】前記記憶装置は少なくとも１つ以上のデータを記憶する物理記憶手段を有し、
前記計算機が前記記憶装置をアクセスする際に利用する論理的な格納位置を前記記憶装置内で実際にデータを記憶する物理記憶手段の記憶位置へ変換する論理－物理位置変換手段を有し、
前記論理位置に対応するデータの前記物理記憶手段における記憶位置を変更するデータの配置変更手段を有し、
前記情報取得手段により取得した情報を利用する前記論理位置に対応するデータの物理記憶位置の変更案を作成する配置変更案作成手段を有することを特徴とする請求項１に記載の記憶装置。

【請求項８】前記配置変更手段を用いて前記配置変更案作成手段により作成されたデータの配置の変更を行なう自動データ再配置手段を有することを特徴とする請求項７に記載の記憶装置。

【請求項９】前記配置変更案作成手段が、前記情報取得手段により取得した情報を基づいて、前記データベース管理システムが前記データベースデータをシーケンシャルにアクセスする際のアクセス場所とアクセス順を判断し、前記シーケンシャルにアクセスされるデータベースデータを前記物理記憶手段上で前記連続アクセス順の前後関係を崩さずに連続した領域に配置することを特徴とする請求項７に記載の記憶装置。

【請求項１０】前記情報取得手段が取得する情報は、前記データベース管理システムが前記スキーマにより定義される同一のデータ構造に属する前記データベースデータをアクセスする際の並列度に関する情報を含み、
前記配置変更案作成手段が、前記取得情報を基に、前記スキーマにより定義される同一のデータ構造に属する前記データベースデータを複数の前記物理記憶手段に配置することを特徴とする請求項７に記載の記憶装置。

【請求項１１】前記配置変更案作成手段が、前記取得情報を基づいて、同時にアクセスされる可能性が高い前記データベースデータの組を検出し、それらを異なる前記物理記憶手段に配置することを特徴とする請求項７に記載の記憶装置。

【請求項１２】前記配置変更案作成手段が、表データと前記表データに対する索引データを抽出し、前記索引が木構造のデータ構造の場合にそれらを前記同時にアクセスされる可能性が高いデータベースデータの組として登録することを特徴とする請求項１１に記載の記憶装置。

【請求項１３】前記データベースに関する情報に、前記データベース管理システムにおける処理の実行履歴に関する情報を含むことを特徴とする請求項１１に記載の記憶装置。

【請求項１４】前記物理記憶手段の稼動情報を取得する物理記憶手段稼動情報取得手段を有し、前記配置変更案作成手段が前記物理記憶手段稼動情報取得手段により取得した情報を利用することを特徴とする請求項１１に記載の記憶装置。

【請求項１５】前記配置変更案作成手段が、前記データベース管理システムにおけるデータの更新処理時に記録するログデータとその他の前記データベースデータを前記同時にアクセスされる可能性が高いデータベースデータの組として登録することを特徴とする請求項１１に記載の記憶装置。

【請求項１６】キャッシュメモリとデータを記憶する物理記憶手段を有し、
前記情報取得手段により取得した情報を利用して前記キャッシュメモリを制御するキャッシュメモリ制御手段を有し、
前記情報取得手段が取得する情報として、前記データベース管理システムが前記計算機上のメモリ上に有するキャッシュの前記データベースデータに対する制御方法とそのキャッシュデータ量に関する情報である計算機キャッシュデータ情報を含むことを特徴とする請求項１に記載の記憶装置。

【請求項１７】前記キャッシュメモリ制御手段が、前記スキーマにより定義されるデータ構造に関して前記計算機上メモリにキャッシュされているデータ量である計算機データ構造キャッシュデータ量と前記データ構造の実

データのデータサイズを比較し、前記比較結果を用いて前記データ構造の実データ記憶位置に記憶されているデータのキャッシュ優先度を決定することを特徴とする請求項１６に記載の記憶装置。

【請求項１８】前記キャッシュメモリ制御手段が、前記計算機データ構造キャッシュデータ量と前記記憶装置における前記データ構造に属するデータに対して前記キャッシュメモリの利用可能量を比較し、前記比較結果を用いて前記データ構造の実データ記憶位置に記憶されているデータのキャッシュ優先度を決定することを特徴とする請求項１６に記載の記憶装置。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明は、データベース管理システムに関する，

【０００２】

【従来の技術】現在、データベース（ＤＢ）を基盤とする多くのアプリケーションが存在し、ＤＢに関する一連の処理・管理を行うソフトウェアであるデータベース管理システム（ＤＢＭＳ）は極めて重要なものとなっている。特に、ＤＢＭＳの処理性能はＤＢを利用するアプリケーションの性能も決定するため、ＤＢＭＳの処理性能の向上は極めて重要である。

【０００３】ＤＢの特徴の１つは、多大な量のデータを扱うことである。そのため、ＤＢＭＳの実行環境の多くにおいては、ＤＢＭＳが実行される計算機に対して大容量の記憶装置を接続し、記憶装置上にＤＢのデータを記憶する。そのため、ＤＢに関する処理を行う場合に、記憶装置に対してアクセスが発生し、記憶装置におけるデータアクセス性能がＤＢＭＳの性能を大きく左右する。そのため、ＤＢＭＳが稼動するシステムにおいて、記憶装置の性能最適化が極めて重要である。

【０００４】文献“Ｏｒａｃｌｅ８ｉパフォーマンスのための設計およびチューニング，リリース８．１”（日本オラクル社，部品番号Ｊ００９２１－０１）の第２０章（文献１）においては、ＲＤＢＭＳであるＯｒａｃｌｅ８ｉにおけるＩ／Ｏのチューニングについて述べられている。その中で、ＲＤＢＭＳの内部動作のチューニングと共に、データの配置のチューニングに関連するものとして、ログファイルは他のデータファイルから分離した物理記憶装置に記憶すること、ストライプ化されたディスクにデータを記憶することによる負荷分散が効果があること、表のデータとそれに対応する索引データは異なる物理記憶装置に記憶すると効果があること、ＲＤＢＭＳとは関係ないデータを異なる物理記憶装置に記憶することが述べられている。

【０００５】米国特許６０３５３０６（文献２）においては、ＤＢＭＳ－ファイルシステム－ボリュームマネージャー－記憶装置間のマッピングを考慮した性能解析ツールに関する技術を開示している。この性能解析ツールは、各レイヤにおけるオブジェクトの稼動状況を画面に表示する。このときに上記のマッピングを考慮し、その各オブジェクトに対応する他レイヤのオブジェクトの稼動状況を示す画面を容易に表示する機能を提供する。また、ボリュームマネージャレイヤのオブジェクトに関して、負荷が高い記憶装置群に記憶されているオブジェクトのうち、２番目に負荷が高いオブジェクトを、もっとも負荷が低い記憶装置群に移動するオブジェクト再配置案を作成する機能を有している。

【０００６】特開平９－２７４５４４号公報（文献３）においては、計算機がアクセスするために利用する論理的記憶装置を実際にデータを記憶する物理記憶装置上に配置する記憶装置において、前記論理的記憶装置の物理記憶装置への配置を動的に変更することにより記憶装置のアクセス性能を向上する技術について開示している。アクセス頻度が高い物理記憶装置に記憶されているデータの一部を前記の配置動的変更機能を用いて他の物理記憶装置に移動することにより、特定の物理記憶装置のアクセス頻度が高くならないようにし、これにより記憶装置を全体としてみたときの性能を向上させる。また、配置動的変更機能による高性能化処理の自動実行方法についても開示している。

【０００７】特開２００１－６７１８７号公報（文献４）においては、計算機がアクセスするために利用する論理的記憶装置を実際にデータを記憶する物理記憶装置上に配置し、前記論理的記憶装置の物理記憶装置への配置を動的に変更する機能を有する記憶装置において、論理的記憶装置の物理記憶装置への配置の変更案を作成する際に物理記憶装置を属性の異なるグループに分割し、それを考慮した配置変更案を作成し、その配置変更案に従って自動的に論理的記憶装置の配置を変更する技術について開示している。配置変更案作成時に、物理記憶装置を属性毎にグループ化し、論理的記憶装置の配置先として、それが有している特徴にあった属性を保持している物理記憶装置のグループに属する物理記憶装置を選択する配置変更案を作成することによりより良好なものを作成する。

【０００８】米国特許５３１７７２７（文献５）においては、ＤＢＭＳの処理の一部あるいは全部を記憶装置側で実施することによりＤＢＭＳの高速化する技術について開示している。記憶装置側でＤＢＭＳ処理を行うため、記憶装置においてデータアクセス特性を把握することが可能であり、データアクセス特性と記憶装置の構成を考慮することにより無駄な物理記憶媒体に対するアクセスを削減や必要なデータの先読みを実施することができ、結果としてＤＢＭＳの性能を向上させることができる。

【０００９】論文“高機能ディスクにおけるアクセスプランを用いたプリフェッチ機構に関する評価”（向井他著，第１１回データ工学ワークショップ（ＤＥＷＳ２０

００）論文集講演番号３Ｂ－３，２０００年７月発行ＣＤ－ＲＯＭ，主催：電子情報通信学会データ工学研究専門委員会）（文献６）では、記憶装置の高機能化によるＤＢＭＳの性能向上について論じている。具体的には，記憶装置に対してアプリケーションレベルの知識としてリレーショナルデータベース管理システム（ＲＤＢＭＳ）における問い合わせ処理の実行時のデータのアクセスプランを与えた場合の効果について述べている。更に、その確認のために、トレースデータを用いたホスト側から発行する先読み指示により前記技術を模した擬似実験を実施している。

【００１０】

【発明が解決しようとする課題】従来の技術には以下のような問題が存在する。

【００１１】文献１で述べられているものは、管理者がデータの配置を決定する際に考慮すべき項目である。現在、１つの記憶装置内に多数の物理記憶装置を有し、多数の計算機により共有されるものが存在する。この記憶装置においては、多くの場合、ホストが認識する論理的記憶装置を実際にデータを記憶する物理記憶装置の適当な領域に割り当てることが行われる。このような記憶装置を利用する場合、人間がすべてを把握することは困難であり、このような記憶装置を含む計算機システム側に何かしらのサポート機能が存在しなければ文献１に述べられている問題点を把握することすら困難となる。また、問題点を把握することができたとしても、計算機システム側にデータの移動機能が存在しない場合には、記憶装置上のデータの再配置のためにデータのバックアップとリストアが必要となり、その処理に多大な労力を必要とする。

【００１２】文献２で述べられている技術では、ボリュームマネージャレイヤにおけるオブジェクトの稼動状況による配置最適化案を作成する機能を実現しているが、記憶装置から更に高いアクセス性能を得ようとする場合にはＤＢＭＳレイヤにおけるデータの特徴を考慮して配置を決定する必要があるがその点の解決方法に関しては何も述べていない。

【００１３】文献３，文献４で述べられている技術に関しては、記憶装置におけるデータ記憶位置の最適化に関する技術である。これらの技術においては記憶装置を利用するアプリケーションが利用するデータに関する特徴としては、アクセス頻度とシーケンシャルアクセス性程度しか考慮していないため、アプリケーションから見た場合に必ずしも最適な配置が実現できるわけではない。例えば、ＤＢＭＳにおける表データとそれに対応する索引データのような同時にアクセスされるデータを同一の物理記憶装置に配置する可能性がある。このとき、その物理記憶装置においてアクセス競合が発生し、記憶装置のアクセス性能が低下する可能性がある。また、文献１から文献４で述べられている技術に関しては、記憶装置におけるキャッシュメモリの利用については特に考慮されていない。

【００１４】文献５，文献６で述べられている技術に関しては、データの記録位置の最適化に関して特に考慮をしていない。そのため、特定の物理記憶装置の負荷が高くなっているために記憶装置の性能が低下し、ＤＢＭＳの性能が悪化している状況の解決手段として有効なものではない。

【００１５】本発明の第一の目的は、ＤＢＭＳが管理するデータを保持する記憶装置において、ＤＢＭＳの処理の特徴を考慮することによりＤＢＭＳに対してより好ましい性能特性を持つ記憶装置を実現することである。この記憶装置を利用することにより、既存のＤＢＭＳを利用したＤＢシステムに対しても、ＤＢＭＳの性能を向上させることができる。

【００１６】本発明の第二の目的は、記憶装置の性能最適化機能を提供することにより記憶装置の性能に関する管理コストを削減することである。これにより、この記憶装置を用いたＤＢシステムのシステム管理コストを削減することができる。

【００１７】

【課題を解決するための手段】ＤＢＭＳに関する情報を記憶装置が取得し、その情報と記憶装置内の物理記憶装置の特性と更に利用可能であれば記憶装置を利用する他のアプリケーションのアクセス頻度に関する情報を考慮する記憶装置の性能最適化処理を記憶装置上で実施する。

【００１８】記憶装置がＤＢＭＳの特性を考慮してＤＢＭＳに良好な性能を得るための手段の第一として、ホストが認識する論理的記憶装置を物理記憶装置の適当な領域に割り当ててデータを記憶する記憶装置における、論理的記憶装置のデータ記憶位置の最適化が存在する。例えば、データ更新時に必ず書き込みが実行される更新ログを、他のデータと異なる物理記憶装置に配置して相互干渉しないようにすることによりＤＢＭＳに対して良好な性能特性を得ることができる。

【００１９】また、同時にアクセスされる可能性が極めて高い表データとそれに対応する索引データを異なる物理記憶装置に配置することによりＤＢＭＳに対して良好な性能特性を得ることができる。更に、ＤＢＭＳに関する情報を利用して、データがシーケンシャルにアクセスされる場合のアクセス順序を予測し、その構造を保持するように物理記憶装置に記憶することによりシーケンシャルアクセス性能を向上可能である。現在、論理的記憶装置の記憶位置を動的に変更する技術は存在し、これを利用することによりデータの最適配置を実現できる。

【００２０】記憶装置がＤＢＭＳの特性を考慮してＤＢＭＳに良好な性能を得るための手段の第二として、ＤＢＭＳにおけるホスト上のキャッシュ動作を考慮したキャッシュメモリ制御が存在する。ＤＢＭＳにおいては利用

頻度の高いデータをホストのメモリ上にキャッシュするが、全てのデータがホストメモリ上に乗ってしまうようなデータに対しては、記憶装置上のキャッシュに保持してもあまり効果はない。

【００２１】また、多くのＤＢＭＳにおいては、ホスト上のキャッシュの破棄データの選択にＬＲＵアルゴリズムを用いている。ホスト上にキャッシュ可能なデータ量と比べてある一定量以下のデータしか記憶装置上のキャッシュに保持できない場合は、読み出しアクセスにより記憶装置上のキャッシュ上に保持された後にキャッシュに乗っている間に再利用される可能性は低く、そのようなデータを記憶装置上のキャッシュに保持することの効果は小さい。このようなデータをキャッシュから優先的に破棄するような制御を記憶装置上で行うことにより、キャッシュ効果の高いものをより多量に記憶装置のキャッシュメモリ上に保持できるようになり、記憶装置のアクセス性能が向上する。

【００２２】

【発明の実施の形態】以下、本発明の実施の形態を説明する。なお、これにより本発明が限定されるものではない。

＜第一の実施の形態＞本実施の形態では、ＤＢＭＳが実行される計算機と記憶装置が接続された計算機システムにおいて、記憶装置がＤＢＭＳに関する情報、記憶装置外におけるデータの記憶位置のマッピングに関する情報を取得し、それらを用いて記憶装置の動作を改善する。記憶装置において、記憶装置内部でデータの記憶位置を動的に変更する機能を有し、取得した情報をもとに好適なデータ再配置案を作成し、データの記憶位置の動的変更機能を用いて、作成したデータ再配置案に従ったデータ配置を実現し、アクセス性能を改善する。また、取得情報をもとにしたデータキャッシュの制御を行いより良いアクセス性能特性が得られるようにする。

【００２３】図１は、本発明の第一の実施の形態における計算機システムの構成図である。本実施の形態における計算機システムは、ＤＢホスト８０ａ，８０ｂ、ホスト情報設定サーバ８２、記憶装置１０から構成される。ＤＢホスト８０ａ，８０ｂ、ホスト情報設定サーバ８２、記憶装置１０はそれぞれが保有するネットワークインターフェイス７８を通してネットワーク７９に接続されている。また、ＤＢホスト８０ａ，８０ｂ、記憶装置１０はそれぞれが保有するＩ／Ｏパスインターフェイス７０からＩ／Ｏパス７１を介してＩ／Ｏパススイッチ７２に接続され、これらを通して記憶装置１０とＤＢホスト８０ａ，８０ｂ間のデータ転送を行う。

【００２４】本実施の形態においては、記憶装置１０とＤＢホスト８０ａ，８０ｂ間のデータ転送を行うＩ／Ｏパス７１とネットワーク７９を異なるものとしているが、例えばｉＳＣＳＩのような計算機と記憶装置間のデータ転送をネットワーク上で実施する技術も開発されており、本実施の形態においてもこの技術を利用してもよい。このとき、記憶装置１０とＤＢホスト８０ａ，８０ｂにおいてＩ／Ｏパスインターフェイス７０が省かれ、計算機システム内からＩ／Ｏパス７１とＩ／Ｏパススイッチ７２が省かれる構成となる。

【００２５】記憶装置１０は、記憶領域を提供するもので、その記憶領域は記憶領域管理単位であるボリュームを用いて外部に提供し、ボリューム内の部分領域に対するアクセスや管理はブロックを単位として実行する。記憶装置１０は、ネットワークインターフェイス７８、Ｉ／Ｏパスインターフェイス７０、記憶装置制御装置１２、ディスクコントローラ１６、物理記憶装置１８から構成され、ネットワークインターフェイス７８、Ｉ／Ｏパスインターフェイス７０、記憶装置制御装置１２、ディスクコントローラ１６はそれぞれ内部バス２０により接続され、ディスクコントローラ１６と物理記憶装置１８は物理記憶装置バス２２により接続される。記憶装置制御装置１２は、ＣＰＵ２４とメモリ２６を有する。

【００２６】メモリ２６上には、記憶装置におけるキャッシュメモリとして利用するデータキャッシュ２８が割り当てられ、記憶装置を制御するためのプログラムである記憶装置制御プログラム４０が記憶される。また、メモリ２６上には、物理記憶装置１８の稼動情報である物理記憶装置稼動情報３２、データキャッシュ２８の管理情報であるデータキャッシュ管理情報３４、ＤＢホスト８０ａ，８０ｂで実行されているＤＢＭＳ１１０ａ，１１０ｂに関する情報であるＤＢＭＳデータ情報３６、記憶装置１０が提供するボリュームを物理的に記憶する物理記憶装置１８上の記憶位置の管理情報であるボリューム物理記憶位置管理情報３８を保持する。

【００２７】図中の記憶装置１０は、複数の物理記憶装置１８を有し、１つのボリュームに属するデータを複数の物理記憶装置１８に分散配置することが可能である。また、データが記憶される物理記憶装置１８上の位置を動的に変更する機能を有する。記憶装置制御プログラム４０は、ディスクコントローラ１６の制御を行うディスクコントローラ制御部４２、データキャッシュ２８の管理を行うキャッシュ制御部４４、記憶装置１０が提供するボリュームを物理的に記憶する物理記憶装置１８上の記憶位置の管理やデータが記憶される物理記憶装置１８上の位置を動的に変更する機能に関する処理を行う物理記憶位置管理・最適化部４６、Ｉ／Ｏパスインターフェイス７０の制御を行うＩ／Ｏパスインターフェイス制御部４８、ネットワークインターフェイス７８の制御を行うネットワークインターフェイス制御部５０を含む。

【００２８】ＤＢホスト８０ａ，８０ｂ、ホスト情報設定サーバ８２においては，それぞれＣＰＵ８４、ネットワークインターフェイス７８、メモリ８８を有し、メモリ８８上にオペレーティングシステム（ＯＳ）１００が記憶・実行されている。

【００２９】ＤＢホスト８０ａ，８０ｂはＩ／Ｏパスインターフェイス７０を有し、記憶装置１０が提供するボリュームに対してアクセスを実行する。ＯＳ１００内にファイルシステム１０４と１つ以上のボリュームからホストが利用する論理的なボリュームである論理ボリュームを作成するボリュームマネージャ１０２と、ファイルシステム１０４やボリュームマネージャ１０２により、ＯＳ１００によりアプリケーションに対して提供されるファイルや論理ローボリュームに記憶されたデータの記録位置等を管理するマッピング情報１０６を有する。ＯＳ１００が認識するボリュームやボリュームマネージャ１０２により提供される論理ボリュームに対して、アプリケーションがそれらのボリュームをファイルと等価なインターフェイスでアクセスするための機構であるローデバイス機構をＯＳ１００が有していても良い。

【００３０】図中の構成ではボリュームマネージャ１０２が存在しているが、本実施の形態においてはボリュームマネージャ１０２における論理ボリュームの構成を変更することはないので、ボリュームマネージャ１０２が存在せずにファイルシステムが記憶装置１０により提供されるボリュームを利用する構成に対しても本実施の形態を当てはめることができる。

【００３１】ＤＢホスト８０ａ，８０ｂのそれぞれのメモリ８８上ではＤＢＭＳ１１０ａ，１１０ｂが記憶・実行され、実行履歴情報１２２が記憶されている。ＤＢＭＳ１１０ａ，１１０ｂは内部にスキーマ情報１１４を有している。図中では、ＤＭＢＳ１１０ａ，１１０ｂが１台のホストに１つのみ動作しているが、後述するように、ＤＢＭＳ１１０ａ，１１０ｂ毎の識別子を用いて管理を行うため、１台のホストにＤＢＭＳが複数動作していても本実施の形態にあてはめることができる。

【００３２】ＤＢホスト８０ａ上ではＤＢＭＳ情報取得・通信プログラム１１８が動作している。一方、ＤＢホスト８０ｂ上ではＤＢＭＳ情報取得・通信プログラム１１８が提供する機能をＤＢＭＳ１１０ｂ中のＤＢＭＳ情報収集・通信部１１６が提供する。

【００３３】ホスト情報設定サーバ８２のメモリ８８上ではホスト情報設定プログラム１３０が記憶・実行される。

【００３４】図２はＤＢホスト８０ａ，８０ｂのＯＳ１００内に記憶されているマッピング情報１０６を示す。マッピング情報１０６中には、ボリュームローデバイス情報５２０、ファイル記憶位置情報５３０と論理ボリューム構成情報５４０が含まれる。

【００３５】ボリュームローデバイス情報５２０中にはＯＳ１００においてローデバイスを指定するための識別子であるローデバイスパス名５２１とそのローデバイスによりアクセスされる記憶装置１０が提供するボリュームあるいは論理ボリュームの識別子であるローデバイスボリューム名５２２の組が含まれる。

【００３６】ファイル記憶位置情報５３０中には、ＯＳ１００においてファイルを指定するための識別子であるファイルパス名５３１とそのファイル中のデータ位置を指定するブロック番号であるファイルブロック番号５３２とそれに対応するデータが記憶されている記憶装置１０が提供するボリュームもしくは論理ボリュームの識別子であるファイル配置ボリューム名５３３とそのボリューム上のデータ記憶位置であるファイル配置ボリュームブロック番号５３４の組が含まれる。

【００３７】論理ボリューム構成情報５４０中にはボリュームマネージャ１０２により提供される論理ボリュームの識別子である論理ボリューム名５４１とその論理ボリューム上のデータの位置を示す論理ボリューム論理ブロック番号５４２とその論理ブロックが記憶されているボリュームの識別子であるボリューム名５０１とボリューム上の記憶位置であるボリューム論理ブロック番号５１２の組が含まれる。マッピング情報１０６を取得するには、ＯＳ１００が提供している管理コマンドの実行や情報提供機構の利用、場合によっては参照可能な管理データの直接解析等を行う必要がある。

【００３８】図３はＤＢＭＳ１１０ａ，１１０ｂ内に記憶されているその内部で定義・管理しているデータその他の管理情報であるスキーマ情報１１４を示す。スキーマ情報１１４には、表のデータ構造や制約条件等の定義情報を保持する表定義情報５５１、索引のデータ構造や対象である表等の定義情報を保持する索引定義情報５５２、利用するログに関する情報であるログ情報５５３、利用する一時表領域に関する情報である一時表領域情報５５４、管理しているデータのデータ記憶位置の管理情報であるデータ記憶位置情報５５５、キャッシュの構成に関する情報であるキャッシュ構成情報５５６とデータをアクセスする際の並列度に関する情報である最大アクセス並列度情報５５７を含む。

【００３９】データ記憶位置情報５５５中には、表、索引、ログ、一時表領域等のデータ構造の識別子であるデータ構造名５６１とそのデータを記憶するファイルまたはローデバイスの識別子であるデータファイルパス名５６２とその中の記憶位置であるファイルブロック番号５６３の組が含まれる。キャッシュ構成情報５５６はＤＢＭＳ１１０ａ，１１０ｂが三種類のキャッシュ管理のグループを定義し、そのグループに対してキャッシュを割り当てている場合を示す。

【００４０】キャッシュ構成情報５５６中には、グループ名５６５とグループ中のデータ構造のデータをホスト上にキャッシュする際の最大データサイズであるキャッシュサイズ５６６とそのグループに所属するデータ構造の識別子の所属データ構造名５６７の組が含まれる。最大アクセス並列度情報５５７には，データ構造名５６１とそのデータ構造にアクセスする際の一般的な場合の最大並列度に関する情報である最大アクセス並列度５６９

の組が含まれる。

【００４１】スキーマ情報１１４を外部から取得するには、管理ビューとして外部に公開されているものをＳＱＬ等のデータ検索言語を用いて取得したり、または、専用の機構を用いて取得したりすることができる。

【００４２】図４はＤＢホスト８０ａ，８０ｂのメモリ８８上に記憶されている実行履歴情報１２２を示す。実行履歴情報１２２中には、ＤＢＭＳ１１０ａ，１１０ｂで実行されたクエリ５７０の履歴が記憶されている。この情報は、ＤＢＭＳ１１０ａ，１１０ｂで作成する。またはＤＢＭＳのフロントエンドプログラムがこの情報を作成する。この場合には、ＤＢＭＳフロントエンドプログラムが存在する計算機に実行履歴情報１２２が記憶されることになる。

【００４３】図５は記憶装置１０内に保持されているボリューム物理記憶位置管理情報３８を示す。ボリューム物理記憶位置管理情報３８中には、データの論理アドレス－物理記憶装置１８における記憶位置のマッピングを管理するボリューム物理記憶位置メイン情報５１０と記憶装置１０内でのボリュームに属するデータの物理記憶位置の変更処理の管理情報であるボリュームデータ移動管理情報５１１が含まれる。

【００４４】ボリューム物理記憶位置メイン情報５１０中には、ボリューム名５０１とそのボリューム上のデータ記憶位置であるボリューム論理ブロック番号５１２とその論理ブロックが記憶されている物理記憶装置１８の識別子である物理記憶装置名５０２と物理記憶装置１８上の記憶位置である物理ブロック番号５１４の組のデータが含まれる。ここで、ボリューム名５０１が“Ｅｍｐｔｙ”であるエントリ５１５は特殊なエントリであり、このエントリには記憶装置１０内の物理記憶装置１８の領域のうち、ボリュームに割り当てられていない領域を示し、この領域に対してデータをコピーすることによりデータの物理記憶位置の動的変更機能を実現する。

【００４５】ボリュームデータ移動管理情報５１１はボリューム名５０１と、そのボリューム内の記憶位置を変更するデータ範囲を示す移動論理ブロック番号７８２と、そのデータが新規に記憶される物理記憶装置１８の識別子とその記憶領域を示す移動先物理記憶装置名７８３と移動先物理ブロック番号７８４、現在のデータコピー元を示すコピーポインタ７８６とデータの再コピーの必要性を管理する差分管理情報７８５の組が含まれる。

【００４６】差分管理情報７８５とコピーポインタ７８６を用いたデータの記憶位置変更処理の概略を以下に示す。差分管理情報７８５はある一定量の領域毎にデータコピーが必要である「１」または不必要「０」を示すデータを保持する。データの記憶位置変更処理開始時に全ての差分管理情報７８５のエントリを１にセットし、コピーポインタ７８６を移動元の先頭にセットする。コピーポインタ７８６にしたがって差分管理情報７８５に１がセットされている領域を順次移動先にデータをコピーし、コピーポインタ７８６を更新していく。差分管理情報７８５で管理される領域をコピーする直前に、その対応するエントリを０にセットする。

【００４７】データコピー中に移動領域内のデータに対する更新が行われた場合、それに対応する差分管理情報７８５のエントリを１にセットする。一度全領域のコピーが完了した段階で差分管理情報７８５内のエントリが全て０になったかを確認し、全て０であればボリューム物理記憶位置メイン情報５１０を更新してデータの記憶位置変更処理は完了する。１のエントリが残っている場合には、再度それに対応する領域をコピーする処理を前記手順で繰り返す。

【００４８】なお、データ記憶位置の動的変更機能の実現方法は他の方式を用いても良い。この場合には、ボリューム物理記憶位置管理情報３８中にはボリュームデータ移動管理情報５１１ではなく他のデータ記憶位置の動的変更機能のための管理情報が含まれることになる。

【００４９】図６に記憶装置１０内に保持されている物理記憶装置稼動情報３２を示す。物理記憶装置稼動情報３２中には、記憶装置１０が提供するボリュームの識別子であるボリューム名５０１とそのボリューム名５０１を持つボリュームのデータを保持する物理記憶装置１８の識別子である物理記憶装置名５０２、そしてボリューム名５０１を持つボリュームが物理記憶装置名５０２を持つ物理記憶装置１８に記憶しているデータをアクセスするための稼動時間のある時刻からの累積値である累積稼働時間５０３、稼動率５９４計算のために前回利用した累積稼働時間５０３の値である旧累積稼動時間５９３とある一定時間内の動作時間の割合を示す稼働率５９４の組と、稼働率５９４計算のために前回累積稼動時間を取得した時刻である前回累積稼動時間取得時刻５９５を含む。

【００５０】ディスクコントローラ制御部４２はディスクコントローラ１６を利用して物理記憶装置１８へのデータアクセスする際の開始時刻と終了時刻を取得し、そのアクセスデータがどのボリュームに対するものかを判断して開始時刻と終了時刻の差分を稼動時間として対応するボリューム名５０１と物理記憶装置名５０２を持つデータの組の累積稼働時間５０３に加算する。

【００５１】物理記憶位置管理・最適化部４６は一定間隔で以下の処理を行う。累積稼働時間５０３と旧累積稼働時間５９３、前回累積稼動時間取得時刻５９５と現データ取得時刻を用いて前回累積稼動時間取得時刻５９５と現データ取得時刻間の稼働率５９４を計算・記憶する。その後、取得した累積稼働時間５０３を旧累積稼働時間５９３に、現データ取得時刻を前回累積稼動時間取得時刻５９５に記憶する。

【００５２】図７に記憶装置１０内に保持されているＤＢＭＳデータ情報３６を示す。ＤＢＭＳデータ情報３６

中には、ＤＢＭＳスキーマ情報７１１、データ構造物理記憶位置情報７１２、ＤＢＭＳ実行履歴情報７１４、ＤＢＭＳデータ構造キャッシュ効果情報７１５を含む。

【００５３】ＤＢＭＳデータ情報３６中に含まれるデータは、ＤＢホスト８０ａ，８０ｂ上に存在するデータを利用する必要があるものが含まれる。記憶装置１０は記憶装置１０の外部に存在する情報をホスト情報設定サーバ８２で動作するホスト情報設定プログラム１３０を利用して取得する。ホスト情報設定プログラム１３０はネットワーク７９を通し、ＤＢホスト８０ａ上で実行され、マッピング情報１０６等必要となる情報の収集処理を実施するＤＢＭＳ情報取得・通信プログラム１１８や，ＤＢホスト８０ｂ上で実行されているＤＢＭＳ１１０ｂ中のＤＢＭＳ情報取得・通信プログラム１１８と等価な機能を実現するＤＢＭＳ情報収集・通信部１１６を利用して必要な情報を収集する。

【００５４】ホスト情報設定プログラム１３０は情報取得後、必要ならば記憶装置１０に情報を設定するためのデータの加工を行い、ネットワーク７９を通して記憶装置１０に転送する。記憶装置１０においては、ネットワークインターフェイス制御部５０が必要な情報が送られてきたことを確認し、物理記憶位置管理・最適化部４６に渡し、必要な加工を行った後にその情報をＤＢＭＳデータ情報３６中の適切な場所に記憶する。

【００５５】ホスト情報設定プログラム１３０は任意のＤＢホスト８０ａ，８０ｂ上で実行されていてもよい。あるいは、物理記憶位置管理・最適化部４６がホスト情報設定プログラム１３０の情報収集機能を有してもよい。これらの場合は、ＤＢホスト８０ａ，８０ｂから情報を転送する際にＩ／Ｏパス７１を通して行ってもよい。この場合、特定の領域に対する書き込みが特定の意味を持つ特殊なボリュームを記憶装置１０はＤＢホスト８０ａ，８０ｂに提供し、そのボリュームに対する書き込みがあった場合にＩ／Ｏパスインターフェイス制御部７０は情報の転送があったと判断し、その情報を物理記憶位置管理・最適化部４６に渡し、必要な加工を行った後にその情報をＤＢＭＳデータ情報３６中の適切な場所に記憶する、等の方式を利用する。

【００５６】情報の収集処理に関しては，記憶装置１０が必要になったときに外部にデータ転送要求を出す方法と、データの変更があるたびに外部から記憶装置１０に変更されたデータを送る方法の２種類ともに利用することができる。

【００５７】図８にＤＢＭＳデータ情報３６中に含まれるＤＢＭＳスキーマ情報７１１を示す。ＤＢＭＳスキーマ情報７１１は、ＤＢＭＳデータ構造情報６２１、ＤＢＭＳデータ記憶位置情報６２２，ＤＢＭＳパーティション化表・索引情報６２３、ＤＢＭＳ索引定義情報６２４、ＤＢＭＳキャッシュ構成情報６２５、ＤＢＭＳホスト情報６２６を含む。

【００５８】ＤＢＭＳデータ構造情報６２１はＤＢＭＳ１１０ａ，１１０ｂで定義されているデータ構造に関する情報で、ＤＢＭＳ１１０ａ，１１０ｂの識別子であるＤＢＭＳ名６３１、ＤＢＭＳ１１０ａ，１１０ｂ内の表・索引・ログ・一時表領域等のデータ構造の識別子であるデータ構造名５６１、データ構造の種別を表すデータ構造種別６４０、データ記憶位置情報から求めることができるデータ構造が利用する総データ量を示すデータ構造データ量６４１、そのデータ構造をアクセスする際の最大並列度に関する情報である最大アクセス並列度５６９の組を保持する。このとき、データ構造によっては最大アクセス並列度５６９の値を持たない。

【００５９】ＤＢＭＳデータ記憶位置情報６２２はＤＢＭＳ名６３１とそのＤＢＭＳにおけるデータ記憶位置管理情報５５５であるデータ記憶位置管理情報６３８の組を保持する。

【００６０】ＤＢＭＳパーティション化表・索引情報６２３は、１つの表や索引をある属性値により幾つかのグループに分割したデータ構造を管理する情報で、パーティション化されたデータ構造が所属するＤＢＭＳ１１０ａ，１１０ｂの識別子であるＤＢＭＳ名６３１と分割化される前のデータ構造の識別子であるパーティション元データ構造名６４３と分割後のデータ構造の識別子であるデータ構造名５６１とその分割条件を保持するパーティション化方法６４４の組を保持する。今後、パーティション化されたデータ構造に関しては、特に断らない限り単純にデータ構造と呼ぶ場合にはパーティション化後のものを指すものとする。

【００６１】ＤＢＭＳ索引定義情報６２４には、ＤＢＭＳ名６３１、索引の識別子である索引名６３５、その索引のデータ形式を示す索引タイプ６３６、その索引がどの表のどの属性に対するものかを示す対応表情報６３７の組を保持する。

【００６２】ＤＢＭＳキャッシュ構成情報６２５は、ＤＢＭＳ１１０ａ，１１０ｂのキャッシュに関する情報であり、ＤＢＭＳ名６３１とＤＢＭＳ１１０ａ，１１０ｂにおけるキャッシュ構成情報５５６の組を保持する。

【００６３】ＤＢＭＳホスト情報６２６は、ＤＢＭＳ名６３１を持つＤＢＭＳ１１０ａ，１１０ｂがどのホスト上で実行されているかを管理するもので、ＤＢＭＳ名６３１とＤＢＭＳ実行ホストの識別子であるホスト名６５１の組を保持する。ＤＢＭＳスキーマ情報７１１中のＤＢＭＳホスト情報６２６以外は、ＤＢＭＳ１１０ａ，１１０ｂが管理しているスキーマ情報１１４の中から必要な情報を取得して作成する。ＤＢＭＳホスト情報６２６はシステム構成情報で管理者が設定するものである。

【００６４】図９にＤＢＭＳデータ情報３６中に含まれるデータ構造物理記憶位置情報７１２を示す。データ構造物理記憶位置情報７１２はＤＢＭＳ１１０ａ，１１０ｂに含まれるデータ構造が記憶装置１０内でどの物理記

憶装置１８のどの領域に記憶されるかを管理するもので、データ構造を特定するＤＢＭＳ名６３１とデータ構造名５６１、その外部からのアクセス領域を示すボリューム名５０１とボリューム論理ブロック番号５１２、その物理記憶装置１８上の記憶位置を示す物理記憶装置名５０２と物理ブロック番号５１４の組を保持する。この情報は、ＤＢＭＳデータ記憶位置情報６２２とマッピング情報１０６を記憶装置１０の外部から取得し、さらにボリューム物理記憶位置メイン情報５１０を参照して、対応する部分を組み合わせることにより作成する。

【００６５】ＤＢＭＳ１１０ａ，１１０ｂ毎にシーケンシャルアクセスの方法が定まっている。データ構造物理記憶位置情報７１２を作成する際に、ＤＢＭＳ名６３１とデータ構造名５６１により特定されるデータ構造毎に、シーケンシャルアクセス時のアクセス順を保持するようにソートしたデータを作成する。ここでは、対象とするＤＢＭＳ１１０ａ，１１０ｂの種類を絞り、あらかじめデータ構造物理記憶位置情報７１２を作成するプログラムがＤＢＭＳ１１０ａ，１１０ｂにおけるシーケンシャルアクセス方法を把握し、シーケンシャルアクセス順でソートされたデータを作成する。

【００６６】本実施の形態のＤＢＭＳ１１０ａ，１１０ｂにおけるシーケンシャルアクセス方法は以下の方法に従うものとする。あるデータ構造のデータをシーケンシャルアクセスする場合に、データ構造が記憶されているデータファイル名５６２とファイルブロック番号５６３を昇順にソートしその順序でアクセスを実行する。その他にシーケンシャルアクセス方法の決定方法としては、データファイルを管理する内部通番とファイルブロック番号５６３の組を昇順にソートした順番にアクセスする方法等が存在し、それらを利用したシーケンシャルアクセス方法の判断を実施してもよい。

【００６７】図１０にＤＢＭＳデータ情報３６中に含まれるクエリ実行同時アクセスデータ構造カウント情報７１４を示す。これは、実行履歴情報１２２をもとに同時にアクセスされるデータ構造の組と実行履歴中に何回その組を同時にアクセスするクエリが実行されたかを示すデータで、ＤＢＭＳ名６３１、同時にアクセスされる可能性のあるデータ構造のデータ構造名５６１の組を示すデータ構造名Ａ７０１とデータ構造名Ｂ７０２、そして、ＤＢＭＳ実行履歴１２２の解析によりそのデータ構造の組がアクセスされたと判断された回数であるカウント値７０３の組で表される。この組はカウント値７０３の値でソートしておく。

【００６８】クエリ実行時同時アクセスデータカウント情報７１４はＤＢＭＳ実行履歴１２２から作成する。最初にクエリ実行時同時アクセスデータカウント情報７１４のエントリを全消去する。ＤＢＭＳ１１０ａ，１１０ｂにおいて定型処理が行われる場合には、まず、その型により分類し、その型の処理が何回実行されたかを確認する。続いてＤＢＭＳ１１０ａ，１１０ｂから型毎のクエリ実行プランを取得する。そのクエリ実行プランにより示される処理手順から同時にアクセスされるデータ構造の組を判別する。

【００６９】そして、クエリ実行時同時アクセスデータカウント情報７１４中のＤＢＭＳ名６３１・データ構造名Ａ７０１・データ構造名Ｂ７０２を参照し、既に対応するデータ構造の組が存在している場合には先に求めたその型の処理回数をカウント値７０３に加算する。既に対応するデータ構造の組が存在していない場合には、新たにエントリを追加してカウント値７０３を先に求めたその型の処理回数にセットする。

【００７０】ＤＢＭＳ１１０ａ，１１０ｂにおいて非定型処理が行われる場合には、１つ１つの実行されたクエリに関してクエリ実行プランを取得し、そのクエリ実行プランにより示される処理手順から同時にアクセスされるデータ構造の組を判別する。そして、クエリ実行時同時アクセスデータカウント情報７１４中のＤＢＭＳ名６３１・データ構造名Ａ７０１・データ構造名Ｂ７０２を参照し、既に対応するデータ構造の組が存在している場合にはカウント値７０３に１を加算する。既に対応するデータ構造の組が存在していない場合には、新たにエントリを追加してカウント値７０３に１をセットする。

【００７１】クエリ実行プランから同時にアクセスされる可能性があるデータ構造の判別は以下のように行う。まず、木構造の索引に対するアクセスが実施される場合には、その木構造索引データと、その索引が対象とする表データが同時にアクセスされると判断する。また、データの更新処理や挿入処理が行われる場合には、ログとその他のデータが同時にアクセスされると判断する。

【００７２】以下はＤＢＭＳ１１０ａ，１１０ｂの特性に依存するが、例えば、クエリ実行プラン作成時にネストループジョイン処理を多段に渡り実行する計画を作成し、それらの多段に渡る処理を同時に実行するＲＤＢＭＳが存在する。このＲＤＢＭＳを利用する場合にはその多段に渡るネストループジョイン処理で利用する表データとその表に対する木構造の索引データは同時にアクセスされると判断できる。

【００７３】このように、クエリ実行計画による同時アクセスデータの判断に関しては、ＤＢＭＳ１１０ａ，１１０ｂの処理特性を把握して判断する必要があるが、ここでは、対象とするＤＢＭＳ１１０ａ，１１０ｂの種類を絞り、クエリ実行時同時アクセスデータカウント情報７１４を作成するプログラムがＤＢＭＳ１１０ａ，１１０ｂ特有の同時アクセスデータ構造の組を把握できる機能を有することを仮定する。

【００７４】実行履歴情報１２２からクエリ実行時同時アクセスデータカウント情報７１４を作成する処理は、記憶装置１０の内部、外部どちらで実行してもよい。記憶装置１０でクエリ実行時同時アクセスデータカウント

情報７１４を作成する場合には、記憶装置１０がネットワーク７９を通してＤＢホスト８０ａ，８０ｂ、あるいは、実行履歴情報１２２がＤＢＭＳフロントエンドプログラムが実行される計算機上に記憶される場合にはその計算機に対して実行履歴情報１２２を記憶装置１０に転送する要求を出し、その情報をネットワーク７９を通して受け取る。

【００７５】その後、前述のクエリ実行時同時アクセスデータカウント情報７１４作成処理を実施する。記憶装置１０の外部で作成する場合は、例えば、ホスト情報設定サーバ８２がＤＢホスト８０ａ，８０ｂ、あるいはＤＢＭＳフロントエンドプログラムが実行される計算機から実行履歴情報１２２を取得し、クエリ実行時同時アクセスデータカウント情報７１４作成処理を実施する。その後、ネットワーク７９を通して作成されたクエリ実行時同時アクセスデータカウント情報７１４を記憶装置１０に転送し、それをＤＢＭＳデータ情報３６中に記憶する。

【００７６】なお、本実施の形態においては、常に実行履歴情報１２２が作成される必要はない。クエリ実行時同時アクセスデータカウント情報７１４作成時に実行履歴情報１２２が存在しないＤＢＭＳ１１０ａ，１１０ｂが利用するデータ構造に関してはそれらを無視してデータを作成する。また、クエリ実行時同時アクセスデータカウント情報７１４は存在しなくてもよい。

【００７７】図１１にＤＢＭＳデータ情報３６に含まれるＤＢＭＳデータ構造キャッシュ効果情報７１５を示す。ＤＢＭＳデータ構造キャッシュ効果情報７１５は記憶装置１０においてデータ構造をデータキャッシュに保持しておくことに効果があるかどうかを判断した結果を保持するもので、データ構造を特定するＤＢＭＳ名６３１とデータ構造名５６１、そのデータ構造がデータキャッシュに保持する効果があるかどうかの判断結果を示すキャッシュ効果情報７３３を保持する。キャッシュ効果情報７３３の値は、管理者が指定する、もしくは、以下に述べる手順に従って求めるものである。

【００７８】図１２に記憶装置１０において指定されたデータ構造のデータをデータキャッシュに保持する効果があるかどうかの判断する処理のフローを示す。判断基準は２種類有り、１つは「指定データ構造のデータ量に比べてホストキャッシュ量が十分に存在するために利用頻度が高いデータの読出しアクセスが実行されないか」で、もう１つは「記憶装置１０のデータキャッシュ量がホストキャッシュ量に比べて小さく、記憶装置１０のデータキャッシュ量で効果がある利用頻度のデータはホストキャッシュに載ってしまい、記憶装置１０から読出されるデータを記憶装置１０でキャッシュしても効果が低い」ことである。

【００７９】ステップ２８０１で処理を開始する。ステップ２８０２で指定データ構造と同じキャッシュ管理のグループに属するデータ構造のデータ量の総和をＤＢＭＳキャッシュ構成情報６２５とＤＢＭＳデータ構造情報６２１を参照して求める。

【００８０】ステップ２８０３で指定データ構造と同じキャッシュ管理のグループにおけるそのグループの単位データ量あたりのホストにおける平均キャッシュ量を前記のグループのデータ総量とＤＢＭＳキャッシュ構成情報６２５中のキャッシュサイズ５６６から求め、その値をあらかじめ定められたキャッシュ効果判断閾値と比較する。その値が閾値以上の場合にはステップ２８０７に進み、閾値未満の場合にはステップ２８０４に進む。この閾値としては概ね０．７程度の値を用いる。

【００８１】ステップ２８０４では記憶装置１０における単位容量あたりの平均キャッシュデータ量を求める。この値は、記憶装置のデータキャッシュ２８の総容量と外部に提供するボリュームの総容量から求めることができ、これらの値はボリューム物理記憶位置管理情報３８やデータキャッシュ管理情報３４を参照することにより求めることができる。

【００８２】ステップ２８０５では、前述の指定データ構造が属するキャッシュ管理のグループにおける単位データ量あたりのホストにおける平均キャッシュ量に対する記憶装置１０における平均キャッシュ量の比率を求め、その値がキャッシュ効果判断閾値未満の場合はステップ２８０７に進み、閾値以上の場合にはステップ２８０６に進む。この閾値としては概ね０．７程度の値を用いる。

【００８３】ステップ２８０６では記憶装置１０においてキャッシュする効果があると判定し、ステップ２８０８に進みキャッシュ効果判定処理を終了する。

【００８４】ステップ２８０７では記憶装置１０においてキャッシュする効果がないと判定し、ステップ２８０８に進みキャッシュ効果判定処理を終了する。

【００８５】記憶装置１０は、データキャッシュをある一定サイズの領域であるセグメントと呼ぶ管理単位を用いて管理する。図１３に記憶装置１０内に保持されているデータキャッシュ管理情報３４を示す。データキャッシュ管理情報３４中には、データキャッシュ２８のセグメントの状態を示すキャッシュセグメント情報７２０とキャッシュセグメントの再利用対象選定に利用するキャッシュセグメント利用管理情報７４０を含む。

【００８６】キャッシュセグメント情報７２０中には、セグメントの識別子であるセグメントＩＤ７２１と、そのセグメントに記憶されているデータ領域を示すボリューム名５１１とボリューム論理ブロック番号５１２、そして、セグメントの状態を示すステータス情報７２２、後述するセグメントの再利用選定管理に利用するリストの情報であるリスト情報７２３を含む。

【００８７】ステータス情報７２２が示すセグメントの状態としては、物理記憶装置１８上にセグメント内のデ

ータと同じデータが記憶されている“ノーマル”、セグメント内にのみ最新のデータが存在する“ダーティ”、セグメント内に有効なデータが存在しない“インバリッド”が存在する。リスト情報７２３には、現在そのセグメントが属するリストの識別子と、そのリストのリンク情報が記憶される。図中では、リストは双方向リンクリストであるとしている。

【００８８】キャッシュセグメント利用管理情報７４０中には、キャッシュセグメントの再利用対象選定に利用する３種類の管理リストである第１ＬＲＵリスト、第２ＬＲＵリスト、再利用ＬＲＵリストの管理情報として、第１ＬＲＵリスト情報７４１、第２ＬＲＵリスト情報７４２、再利用ＬＲＵリスト情報７４３が記憶される。

【００８９】第１ＬＲＵリスト情報７４１、第２ＬＲＵリスト情報７４２、再利用ＬＲＵリスト情報７４３は、それぞれリストの先頭であるＭＲＵセグメントＩＤ、最後尾であるＬＲＵセグメントＩＤ、そのリストに属するセグメント数を記憶する。この３種類の管理リストはホストからのアクセス要求の処理にかかわるもので、アクセス要求処理の説明時に同時に行う。

【００９０】ホストからのデータアクセス要求があったときの処理を説明する。

【００９１】図１４に記憶装置１０がホストからデータの読出し要求を受け取ったときの処理フローを示す。ステップ２９０１で、Ｉ／Ｏパスインターフェイス７０はホストからのデータ読出し要求を受け、Ｉ／Ｏパスインターフェイス制御部４８がその要求を認識する。

【００９２】ステップ２９０２でキャッシュ制御部４４は読出し要求があったデータがデータキャッシュ２８上に存在するかデータキャッシュ管理情報３４を参照して確認する。存在する場合にはステップ２９０５に進み、存在しない場合にはステップ２９０３に進む。

【００９３】ステップ２９０３で、キャッシュ制御部４４は読出し要求があったデータを保持するキャッシュ領域を確保する。データを保持するキャッシュセグメントとして、ステータス情報がノーマルのもののうち、再利用ＬＲＵリストのＬＲＵ（Ｌｅａｓｔ　Ｒｅｃｅｎｔｌｙ　Ｕｓｅｄ：最も昔に使われた）側に存在するものを必要数取得し、再利用ＬＲＵリストから削除する。そして、再利用ＬＲＵリスト情報７４３をそれに合わせて更新する。また、キャッシュセグメント情報７２０中のボリューム名５１１とボリューム論理ブロック番号を記憶するデータのものに変更し、ステータス情報７２２をインバリッドに設定する。

【００９４】ステップ２９０４でディスクコントローラ制御部４２は読出し要求があったデータを物理記憶装置１８から読み出す処理を実施し、その完了を待つ。読出し完了後、キャッシュセグメント情報７２０中の対応するステータス情報７２２をノーマルに設定し、ステップ２９０６に進む。

【００９５】ステップ２９０５でキャッシュ制御部４４はデータ読出し要求のあったデータを保持するセグメントをその管理のためにリンクされている管理リストから削除する。

【００９６】ステップ２９０６でＩ／Ｏパスインターフェイス管理部４８はデータ読出し要求のあったデータをセグメントからＩ／Ｏパスインターフェイス７０を利用してホストに転送し、ホストとの処理を完了する。

【００９７】ステップ２９０７でキャッシュ制御部４４はアクセス先のデータの内容に従い、データ読出し要求のあったデータを保持するセグメントを適当な管理リストに繋ぐ処理を行う。この処理の詳細は後述する。

【００９８】ステップ２９０８でホストからの読出し要求を受けっとった時の処理を終了する。

【００９９】図１５に記憶装置１０がホストからデータの書き込み要求を受け取ったときの処理フローを示す。ステップ２９３１で、Ｉ／Ｏパスインターフェイス７０はホストからのデータ書き込み要求を受け、Ｉ／Ｏパスインターフェイス制御部４８がその要求を認識する。

【０１００】ステップ２９３２でキャッシュ制御部４４は読出し要求があったデータを保持するセグメントがデータキャッシュ２８上に存在するかデータキャッシュ管理情報３４を参照して確認する。存在する場合にはステップ２９３４に進み、存在しない場合にはステップ２９３３に進む。

【０１０１】ステップ２９３３で、キャッシュ制御部４４は書き込み要求があったデータを保持するキャッシュ領域を確保する。データを保持するキャッシュセグメントとして、ステータス情報がノーマルのもののうち、再利用ＬＲＵリストのＬＲＵ側に存在するものを必要数取得し、再利用ＬＲＵリストから削除する。そして、再利用ＬＲＵリスト情報７４３をそれに合わせて更新する。また、キャッシュセグメント情報７２０中のボリューム名５１１とボリューム論理ブロック番号を記憶するデータのものに変更し、ステータス情報７２２をインバリッドに設定する。

【０１０２】ステップ２９３４でキャッシュ制御部４４はデータ書込み要求のあったデータを保持するセグメントをその管理のためにリンクされている管理リストから削除する。

【０１０３】ステップ２９３５でＩ／Ｏパスインターフェイス管理部４８はデータ書込み要求のあったデータをキャッシュセグメントに書込み、キャッシュセグメント情報７２０中の対応するステータス情報７２２をダーティに設定し、ホストとの処理を完了する。

【０１０４】ステップ２９３６でキャッシュ制御部４４はアクセス先のデータの内容に従い、データ書込み要求のあったデータを保持するセグメントを適当な管理リストに繋ぐ処理を行う。この処理の詳細は後述する。

【０１０５】ステップ２９３７でホストからの書込み要

求を受けっとった時の処理を終了する。

【０１０６】図１６にキャッシュ制御部４４が実行するアクセス先のデータの内容に従い、アクセス要求のあったデータを保持するセグメントを適当な管理リストに繋ぐ処理のフローを示す。この処理において、記憶装置１０におけるキャッシュ効果がないと判断されるデータを保持するキャッシュセグメントを管理リスト中の再利用されやすい場所に繋ぐことによりキャッシュ効果がないと判断されるものがデータキャッシュ２８上に載っている時間を短くし、他のデータのキャッシュ効果を高めることを行う。

【０１０７】ステップ２９６１においてアクセス先のデータの内容に従い、アクセス要求のあったデータを保持するセグメントを適当な管理リストに繋ぐ処理を開始する。

【０１０８】ステップ２９６２において、アクセス先データのキャッシュ効果の確認を行う。データ構造物理記憶位置情報７１２を参照してアクセス先データが属するＤＢＭＳ１１０ａ，１１０ｂとそのデータ構造の識別子であるＤＢＭＳ名６３１とデータ構造名５６１を求める。データ構造物理記憶位置情報７１２に対応部分がない場合にはキャッシュ効果があると判断する。

【０１０９】続いて、ＤＢＭＳデータ構造キャッシュ効果情報７１５を参照し、既に求めたＤＢＭＳ名６３１とデータ構造名５６１に対応するキャッシュ効果情報７３３を参照し、アクセス先データにキャッシュ効果があるかないかを求める。なお、キャッシュ効果情報７３３中に対応するエントリがない場合、キャッシュ効果があると判断する。キャッシュ効果があると判断された場合にはステップ２９６３に進み、ないと判断された場合にはステップ２９６６に進む。

【０１１０】ステップ２９６３でアクセス先データを保持するキャッシュセグメントを第１ＬＲＵリストのＭＲＵ（Ｍｏｓｔ　Ｒｅｃｅｎｔｌｙ　Ｕｓｅｄ：最も最近使われた）側にリンクし、それにあわせて第１ＬＲＵリスト情報７４１を更新する。

【０１１１】ステップ２９６４で第１ＬＲＵリストにリンクされているセグメント数を第１ＬＲＵリスト情報７４１を参照して確認し、その値が事前に定めてある閾値を超えているか確認する。そのセグメント数が閾値未満の場合にはステップ２９７０に進み処理を完了する。閾値以上の場合にはステップ２９６５に進む。

【０１１２】ステップ２９６５で第１ＬＲＵリストのセグメント数が閾値未満になるように第１ＬＲＵの最もＬＲＵ側に存在するセグメントを第２ＬＲＵリストのＭＲＵ側にリンクし直す処理を行い、それに合わせて第１ＬＲＵリスト情報７４１と第２ＬＲＵリスト情報７４２を更新し、ステップ２９６７に進む。

【０１１３】ステップ２９６６でアクセス先データを保持するキャッシュセグメントを第２ＬＲＵリストのＭＲＵ側にリンクし、それに合わせて第２ＬＲＵリスト情報７４２を更新し、ステップ２９６７に進む。

【０１１４】ステップ２９６７で第２ＬＲＵリストにリンクされているセグメント数を第２ＬＲＵリスト情報７４２を参照して確認し、その値が事前に定めてある閾値を超えているか確認する。そのセグメント数が閾値未満の場合にはステップ２９７０に進み処理を完了する。閾値以上の場合にはステップ２９６８に進む。

【０１１５】ステップ２９６８で第２ＬＲＵリストのセグメント数が閾値未満になるように第２ＬＲＵの最もＬＲＵ側に存在するセグメントを再利用ＬＲＵリストのＭＲＵ側にリンクし直す処理を行い、それに合わせて第２ＬＲＵリスト情報７４２と再利用ＬＲＵリスト情報７４３を更新する。

【０１１６】ステップ２９６９で、ステップ２９６８において第２ＬＲＵリストから再利用ＬＲＵリストにリンクしなおされたセグメントに関して、キャッシュセグメント情報７２０中のステータス情報７２２を参照して、その値がダーティであるもののデータを物理記憶装置１８に書き出す処理をディスクコントローラ制御部４２に対して要求し、その完了を待つ。書き出し処理完了後、キャッシュセグメント情報７２０中の対応するステータス情報７２２をノーマルに変更し、ステップ２９７０に進む。

【０１１７】ステップ２９７０で処理を終了する。

【０１１８】図１７に物理記憶位置管理・最適化部４２が実施するデータ再配置処理の処理フローを示す。ここで、管理者の指示により処理を開始するモードと、あらかじめ設定されている時刻に自動的にデータ再配置案作成処理を実施し、その後に作成されたデータ再配置案を実現するためにデータ移動を自動実行するデータ再配置自動実行モードの２種類を考える。

【０１１９】後述するように、複数の異なった種類のデータ配置解析・データ再配置案作成処理を実行可能であり、処理すべき種類の指定をして処理を開始する。また、処理にパラメータが必要な場合は併せてそれが指定されているものとする。これらは、管理者が処理を指示する場合にはそのときに一緒に指示を出し、データ再配置自動実行モードの場合には処理する種類や必要なパラメータを事前に設定しておく。

【０１２０】ステップ２００１でデータ再配置処理を開始する。このとき、データ配置解析・データ再配置案作成処理として何を実行するか指定する。また、必要であればパラメータを指定する。

【０１２１】ステップ２００２でデータ再配置処理に必要なＤＢＭＳデータ情報３６を前述した方法で取得し記憶する。なお、このデータ収集は、ステップ２００１の処理開始とは無関係にあらかじめ実行しておくこともできる。この場合には、情報を取得した時点から現在まで情報に変更がないかどうかをこのステップで確認する。

【０１２２】ステップ２００３では、ワーク領域を確保し、その初期化を行う。ワーク領域としては、図１８に示すデータ再配置ワーク情報６７０と図１９に示す移動プラン情報７５０を利用する。データ再配置ワーク情報６７０と移動プラン情報７５０の詳細とその初期データ作成方法は後述する。

【０１２３】ステップ２００４でデータ配置の解析・再配置案の作成処理を実行する。後述するように、データ配置の解析・再配置案作成処理は複数の観点による異なったものが存在し、このステップではステップ２００１で指定された処理を実行する。またステップ２００１でパラメータを受け取った場合には、それを実行する処理に与える。

【０１２４】ステップ２００５ではステップ２００４のデータ再配置案作成処理が成功したかどうか確認する。成功した場合にはステップ２００６に進む。失敗した場合にはステップ２０１０に進み、管理者にデータ再配置案作成が失敗したことを通知し、ステップ２０１１に進み処理を完了する。

【０１２５】ステップ２００６では、現在データ再配置自動実行モードにより処理を実行しているか確認する。自動実行モードにより処理を実行している場合にはステップ２００９に進む。そうでない場合には、ステップ２００７に進む。

【０１２６】ステップ２００７では、ステップ２００４で作成されたデータ再配置案を管理者に提示する。この提示を受けて管理者はデータ再配置案に問題がないか判断する。ステップ２００８では、データの再配置を続行するか否かを管理者から指示を受ける。続行する場合にはステップ２００９に進み、そうでない場合にはステップ２０１１に進み処理を完了する。

【０１２７】ステップ２００９では、ステップ２００４で作成されたデータの再配置案を基にデータの再配置処理を実行する。このとき、移動プラン情報７５０中の移動順序７６１で示される順に指定されたボリュームの領域を指定された物理記憶装置１８内の領域へデータの移動を実行する。移動処理機能の実現方法は前述した通りである。

【０１２８】ステップ２０１０でデータ再配置処理は完了である。

【０１２９】図１８はステップ２００３において作成する情報であるデータ再配置ワーク情報６７０を示す。データ再配置ワーク情報６７０はデータ移動可能領域を保持する空き領域情報６８０とデータ構造物理記憶位置情報７１２のコピーを保持する。空き領域情報６８０は、データ移動可能領域を示す物理記憶装置名５０２と物理ブロック番号５１４の組を保持する。

【０１３０】データの初期化は以下の方法で実行する。空き領域情報６８０はボリューム物理記憶位置メイン情報５１０中のボリューム名５０１が“Ｅｍｐｔｙ”である領域を集めることにより初期化する。データ構造物理記憶位置情報７１２はＤＢＭＳデータ情報３６に存在するデータをそのままコピーする。データ再配置案作成時にこれらのデータの値を変更するため、データ構造物理記憶位置情報７１２は必ずコピーを作成する。

【０１３１】図１９はステップ２００４で実行されるデータ配置解析・データ再配置案作成処理により作成されるデータ移動案を格納する移動プラン情報７５０を示す。移動プラン情報７５０は、移動指示の実行順序を示す移動順序７６１、移動するデータを持つボリュームとそのデータ領域示す移動ボリューム名７６８と移動ボリューム論理ブロック番号７６９、そのデータの移動先の物理記憶装置とその記憶領域を示す移動先物理記憶装置名７７１と移動先物理ブロック番号７７２の組を保持する。この情報に関しては、何もデータを持たないように初期化する。

【０１３２】ステップ２００４で実行されるデータ配置解析・データ再配置案作成処理について説明する。前述のように、この処理には幾つかの種類が存在する。全ての処理に共通するのは逐次的にデータ再配置のためのデータ移動案を作成することである。そのため、データ移動の順番には意味があり、移動プラン情報７５０中の移動順序７６１にその順番を保持し、その順序どおりにデータ移動を行うことによりデータの再配置を実施する。

【０１３３】また、逐次処理のため、移動後のデータ配置をもとに次のデータの移動方法を決定する必要がある。そこで、データ移動案を作成するたびにデータ再配置ワーク情報６７０をデータ移動後の配置に更新する。

【０１３４】データ再配置案作成時のデータ移動案の作成は以下のように行う。移動したいデータ量以上の連続した移動可能領域をデータ再配置ワーク情報６７０中の情報から把握し、その中の領域を適当に選択し、設定条件や後述する制約を満たすかどうか確認をする。もし、それらを満たす場合にはそこを移動先として設定する。それらを満たさない場合には他の領域を選択し、再度それらを満たすかどうか確認をする。

【０１３５】以下、設定条件と制約を満たす領域を発見するか、全ての移動したいデータ量以上の連続した移動可能領域が設定条件や制約を満たさないことを確認するまで処理を繰り返す。もし、全ての領域で設定条件や制約を満たさない場合にはデータ移動案の作成に失敗として終了する。

【０１３６】このときに重要なのは移動後のデータ配置において、問題となる配置を行わないことである。特にＲＤＢＭＳにおいては、特定のデータに関してはアクセスが同時に行われる可能性が高く、それらを異なる物理記憶装置１８上に配置する必要がある。そこで、以下で説明する全てのデータの移動案を作成する場合には、移動するデータに含まれるデータ構造と、移動先に含まれるデータ構造を調べ、ログとその他のデータ，一時表領

域とその他のデータ、表データとそれに対して作成された木構造の索引データがデータの移動後に同じ物理記憶装置１８に配置されるかどうかを確認し、配置される場合には、その配置案は利用不可能と判断する。

【０１３７】なお、あるデータ構造がどの物理記憶装置１８の領域に記憶されているか、また逆に、ある物理記憶装置１８上の領域に記憶されるデータがどのデータ構造に対応するかは、データ再配置ワーク情報６７０中のデータ構造物理記憶位置情報７１２により把握可能である。

【０１３８】図２０に第１のデータ配置解析・データ再配置案作成処理である、物理記憶装置稼動情報３２を基にした同時アクセス実行データ構造を分離するためのデータ再配置案作成処理の処理フローを示す。本処理においては、物理記憶装置１８の稼動率が閾値を超えたものはディスクネック状態にあると判断してそれを解消するデータの移動案を作成する。本処理は、実測値に基づいて問題点を把握し、それを解決する方法を見つけるため、より精度の高いデータ再配置案を作成すると考えられ、データ再配置自動実行モードで最も利用しやすいものである。

【０１３９】ステップ２１０１で処理を開始する。本処理を開始するにあたっては、どの期間の稼動率を参照するかを指定する。

【０１４０】ステップ２１０２では、物理記憶装置１８の識別子と指定期間における物理記憶装置１８の稼動率の組を記憶するワーク領域を取得し、物理記憶装置稼動情報３２を参照してその情報を設定し、物理記憶装置１８の稼動率で降順にソートする。物理記憶装置稼動情報３２中では、同じ物理記憶装置１８中に記憶されているデータであっても異なるボリュームのものは分離して稼動率を取得しているため、それらの総和として物理記憶装置１８の稼動率を求める必要がある。

【０１４１】ステップ２１０３では、ステップ２１０２のソート結果をもとに物理記憶装置１８の稼動率が閾値を超えているもののリストである過負荷確認リストを作成する。このリスト中のエントリに関しても稼動率が降順になるような順序を保つようにする。

【０１４２】ステップ２１０４では、過負荷確認リスト中にエントリが存在するか確認する。エントリが存在しない場合には、もう過負荷状態の物理記憶装置１８が存在しないものとしてステップ２１０５に進みデータ再配置案作成処理成功として処理を終了する。エントリが存在する場合には、ステップ２１０６に進む。

【０１４３】ステップ２１０６では、過負荷確認リスト中の最も物理記憶装置１８の稼動率が高いものを再配置対象の物理記憶装置１８として選択する。

【０１４４】ステップ２１０７では、再配置対象となった物理記憶装置１８内部のボリュームとその稼動率のリストを物理記憶装置稼動情報３２を参照して作成し、稼動率で降順にソートする。

【０１４５】ステップ２１０８では、リスト中のあるボリュームの稼動率があらかじめ定められた閾値を超過しているかどうか確認する。全てのボリュームの稼動率が閾値を超えていない場合には、ステップ２１１３に進み、あるボリュームの稼動率がその閾値を超えている場合には、ステップ２１０９に進む。

【０１４６】ステップ２１０９においては、稼動率が閾値を超えているボリュームに関して、確認対象の物理記憶装置１８中に同時にアクセスされる可能性があるデータの組、すなわち、ログとその他のデータ，一時表領域とその他のデータ、表データとそれに対して作成された木構造の索引データがあるそのボリューム内部に記憶されているかどうか発見する処理を行う。

【０１４７】ステップ２１１０では、ステップ２１０９における結果を確認し、同時アクセスデータ構造の組が存在する場合にはステップ２１１１に進む。同時アクセスデータ構造の組が存在しない場合には、ステップ２１１２に進む。

【０１４８】ステップ２１１１においては、同時アクセスデータ構造の組に属するデータを異なる物理記憶装置１８に記憶するためのデータ移動案を作成し、ステップ２１１４に進む。

【０１４９】ステップ２１１２においては、現在確認対象となっているボリューム内のデータを論理ブロック番号に従って２分割し、その片方を他の物理記憶装置１８へ移動するデータ移動案を作成し、ステップ２１１４に進む。

【０１５０】ステップ２１１３においては、現在確認対象になっている物理記憶装置１８の稼動率が閾値を下回るまで、稼動率が高いボリュームから順に、その物理記憶装置１８に記憶されているボリュームを構成するデータ全体を他の物理記憶装置１８に移動するデータ移動案を作成し、ステップ２１１４に進む。

【０１５１】ステップ２１１１，２１１２，２１１３のデータ移動先を発見する際に、移動後の移動先の記憶装置の稼動率を予測する。物理記憶装置１８毎の性能差が既知の場合にはその補正を行った移動データを含む記憶装置１８上のボリュームの稼動率分、未知の場合には補正を行わない移動データを含む記憶装置１８上のボリュームの稼動率分、データ移動により移動先の物理記憶装置１８の稼動率が上昇すると考え，加算後の値が閾値を越えないような場所へのデータの移動案を作成する。

【０１５２】稼動率の加算分に関して、移動データ量の比率を考慮しても良いが、ここではデータ中のアクセスの偏りを考慮して移動データに全てのアクセスが集中したと考えた判断を行う。

【０１５３】ステップ２１１４では、データ移動案の作成に成功したかどうかを確認し、失敗した場合にはステップ２１１７に進みデータの再配置案作成処理失敗とし

て処理を終了する。成功した場合にはステップ２１１５に進む。

【０１５４】ステップ２１１５では作成したデータ移動案を移動プラン情報７５０に追加し、ステップ２１１６に進む。ステップ２１１６ではデータ再配置ワーク情報６７０を作成したデータ移動案に従って修正し、移動先記憶装置１８のステップ２１０２で作成した物理記憶装置１８毎の稼動情報の値を前述の移動後の稼動率判断値に修正する。その後、現在の確認対象の物理記憶装置１８を過負荷確認リストから削除し、ステップ２１０４に戻り次の確認を行う。

【０１５５】次に第２のデータ配置解析・データ再配置案作成処理である、クエリ実行同時アクセスデータ構造カウント情報７１４を用いた同時アクセス実行データ構造を分離するためのデータ再配置案作成処理の処理フローを示す。本処理においては、クエリ実行同時アクセスデータ構造カウント情報７１４から同時にアクセスされるデータの組を取得し、それらを異なる物理記憶装置１８に配置するデータ再配置案を作成する。

【０１５６】図２１にクエリ実行時同時アクセスデータカウント情報７１４を利用する同時アクセス実行データ構造を分離するためのデータ再配置案作成処理処理フローを示す。ステップ２２０１で処理を開始する。ステップ２２０２において、カウント値７０３の全エントリの総和に対してカウント値７０３の値が一定割合以上のデータ構造とその所属するＤＢＭＳ１１０ａ，１１０ｂの組を求め、それらを確認リストとして記憶する。

【０１５７】ステップ２２０３でステップ２２０２で求めた確認リスト中に含まれるデータ構造の組に関して、それらを異なる物理記憶装置１８に記憶するデータ再配置案を作成し、ステップ２２０４に進む。なお、ステップ２２０３の処理に関しては、図２２を用いて後で説明する。ステップ２２０４では、ステップ２２０３においてデータ再配置案の作成に成功したかどうかを確認し、成功した場合にはステップ２２０５に進みデータ再配置案作成処理成功として処理を終了し、失敗した場合にはステップ２２０６に進みデータ再配置案作成処理失敗として処理を終了する。

【０１５８】図２２に指定されたデータ構造とそのデータ構造と同時にアクセスされる可能性が高いデータ構造の組を分離するデータ再配置案を作成する処理のフローを示す。本処理を開始するときには、データ構造名と物理記憶装置１８から分離するデータ構造名の組のリストである確認リストを与える。

【０１５９】ステップ２３０１で処理を開始する。ステップ２３０３で確認リスト中にエントリが存在するか確認し、存在しない場合にはステップ２３０４に進みデータ再配置案作成処理成功として処理を終了する。存在する場合にはステップ２３０５に進む。

【０１６０】ステップ２３０５においては、確認リストから１つ確認対象データ構造名とその所属ＤＢＭＳ名の組とその分離データ構造名とその所属ＤＢＭＳ名の組の組を取得し、ステップ２３０６に進む。

【０１６１】ステップ２３０６においては、確認対象データ構造とその分離するデータ構造が同一の物理記憶装置上に記憶されているかどうかの確認を行う。この確認はデータ再配置ワーク情報６７０中のデータ構造物理記憶位置情報７１２を参照することにより可能である。両データ構造が全て異なる物理記憶装置上に存在する場合にはステップ２３１２に進み、ある物理記憶装置上に両データ構造が存在する場合にはステップ２３０７に進む。

【０１６２】ステップ２３０７においては、同一の物理記憶装置上に両データ構造が存在する部分に関してそれを分離するデータ移動案を作成する。ステップ２３０８においては、そのデータ移動案作成が成功したかどうか確認し、成功した場合にはステップ２３１０に進み、失敗した場合にはステップ２３０９に進みデータ再配置案作成処理失敗として処理を終了する。

【０１６３】ステップ２３１０においては、作成されたデータ移動案を移動プラン情報７５０に記憶する。ステップ２３１１においては、作成されたデータ移動案に従ってデータ再配置ワーク情報６７０を更新し、ステップ２３１２に進む。

【０１６４】ステップ２３１２においては、確認リストから現在確認対象となっているデータ構造の組のエントリを削除し、２３０３に進む。

【０１６５】図２３に第３のデータ配置解析・データ再配置案作成処理である、データ構造の定義を基にした同時アクセス実行データ構造を分離するためのデータ再配置案作成処理の処理フローを示す。本処理においては、同時にアクセスされる可能性が高い、ログとその他のデータ、一時表領域とその他のデータ、表データとそれに対して作成された木構造の索引データが同一物理記憶装置１８上に記憶されている部分が存在しないか確認をし，そのような部分が存在する場合にはそれを解決するデータ再配置案を作成する。

【０１６６】ステップ２４０１で処理を開始する。ステップ２４０２では、ＤＢＭＳデータ構造情報６２１を参照して全てのログであるデータ構造名５６１とそれを利用するＤＢＭＳ１１０ａ，１１０ｂのＤＢＭＳ名６３１の組を取得する。そして、そのデータ構造名とログ以外のデータを分離するデータ構造とする確認リストを作成し、ステップ２４０３に進む。　ステップ２４０３ではステップ２４０２で作成した確認リストを用いてステップ２３０１から開始されるデータ構造分離のためのデータ再配置案作成処理を実行する。

【０１６７】ステップ２４０４ではステップ２４０３におけるデータ再配置案作成処理が成功したか確認をし、成功した場合にはステップ２４０５に進む。失敗した場

合にはステップ２４１２に進みデータ再配置案作成処理失敗として処理を終了する。

【０１６８】ステップ２４０５では、ＤＢＭＳデータ構造情報６２１を参照して全ての一時表領域であるデータ構造名５６１とそれを利用するＤＢＭＳ１１０ａ，１１０ｂのＤＢＭＳ名６３１の組を取得する。そして、そのデータ構造と一時表領域以外のデータを分離するデータ構造とする確認リストを作成し、ステップ２４０６に進む。　ステップ２４０６ではステップ２４０５で作成した確認リストを用いてステップ２３０１から開始されるデータ構造分離のためのデータ再配置案作成処理を実行する。

【０１６９】ステップ２４０７ではステップ２４０６におけるデータ再配置案作成処理が成功したか確認をし、成功した場合にはステップ２４０８に進む。失敗した場合にはステップ２４１２に進みデータ再配置案作成処理失敗として処理を終了する。

【０１７０】ステップ２４０８では、ＤＢＭＳ索引定義情報６２４を参照して全ての木構造索引の索引名６３５とそれを利用するＤＢＭＳ１１０ａ，１１０ｂのＤＢＭＳ名６３１の組とそれに対応する表のデータ構造名とそれを利用するＤＢＭＳ１１０ａ，１１０ｂのＤＢＭＳ名６３１の組を対応表情報６３７から取得する。そして、それらの索引と表に関するデータを組とする確認リストを作成し、ステップ２４０９に進む。　ステップ２４０９ではステップ２４０８で作成した確認リストを用いてステップ２３０１から開始されるデータ構造分離のためのデータ再配置案作成処理を実行する。ステップ２４１０ではステップ２４０９におけるデータ再配置案作成処理が成功したか確認をし、成功した場合にはステップ２４１１に進み、データ再配置案作成処理成功として処理を終了する。失敗した場合にはステップ２４１２に進みデータ再配置案作成処理失敗として処理を終了する。

【０１７１】図２４に第４のデータ配置解析・データ再配置案作成処理である、特定の表や索引の同一データ構造に対するアクセス並列度を考慮したデータ再配置案作成処理の処理フローを示す。この処理は、ランダムアクセス実行時の処理の並列度を考慮してディスクネックの軽減を図るためにデータの再配置を行うものである。この処理を実行する際には、データ再配置の確認対象とするデータ構造をＤＢＭＳ名６３１とデータ構造名５６１の組として指定する。

【０１７２】ステップ２５０１で処理を開始する。ステップ２５０２において、指定されたデータ構造の物理記憶装置上に割り当てられた記憶領域利用総量を求める。この値は、ＤＢＭＳデータ構造情報６２１中のデータ構造データ量６４１を参照することにより求める。

【０１７３】ステップ２５０３においては、ＤＢＭＳデータ構造情報６２１を参照して指定データ構造における最大アクセス並列度５６９を取得する。

【０１７４】ステップ２５０４において、ステップ２５０２で求めた指定データ構造の記憶領域利用総量をステップ２５０３で求めた最大アクセス並列度５６９で割った値を、指定データ構造の１つの物理記憶装置１８上への割り当てを許可する最大量として求める。この制約により、特定の物理記憶装置１８に偏ることなく最大アクセス並列度５６９以上の台数の物理記憶装置１８に指定データ構造が分散して記憶されることになり、最大アクセス並列度５６９による並列度でランダムアクセスが実行されてもディスクネックになりにくい状況となる。なお、割り当て許可最大量の値は、実際のアクセス特性を考慮してこの方法で求めた値から更に増減させても構わない。

【０１７５】ステップ２５０５において、指定データ構造のデータがステップ２５０４で求めた最大量を超えて１つの物理記憶装置１８上に割り当てられているものが存在するかデータ再配置ワーク情報６７０を用いて確認し、そのようなものが存在しない場合にはステップ２５０９に進み、データ再配置案作成処理成功として処理を終了する。存在する場合にはステップ２５０６に進む。

【０１７６】ステップ２５０６においては、ステップ２５０４で求めた最大量を超えて１つの物理記憶装置１８上に割り当てられている部分を解消するデータ移動案を作成する。このとき、移動案作成に考慮するデータ移動量は指定データ構造の現在の物理記憶装置１８上への割り当て量のステップ２５０４で求めた最大量からの超過分以上である必要がある。また、移動先物理記憶装置１８においても、移動後にステップ２５０４で求めた最大量を超過しないようにする必要がある。

【０１７７】ステップ２５０７においては、ステップ２５０６のデータ移動案作成処理が成功したか確認をする。成功した場合にはステップ２５０８に進む。失敗した場合にはステップ２５１０に進み、データ再配置案作成処理失敗として処理を終了する。

【０１７８】ステップ２５０８においては作成したデータ移動案を移動プラン情報７５０に記憶し、ステップ２５０９に進みデータ再配置案作成処理成功として処理を終了する。

【０１７９】図２５に第５のデータ配置解析・データ再配置案作成処理である、特定の表データに対するシーケンシャルアクセス時のディスクネックを解消するデータ再配置案作成処理の処理フローを示す。この処理を実行する際には、データ再配置の確認対象とする表をＤＢＭＳ名６３１とデータ構造名５６１の組として指定する。

【０１８０】前述のように、対象とするＤＢＭＳ１１０ａ，１１０ｂの種類が絞られるが、データ構造物理記憶位置情報７１２はシーケンシャルアクセス順にソートされてデータを記憶しているため、シーケンシャルアクセス方法は既知である。

【０１８１】また、並列にシーケンシャルアクセスを実

行する場合に、その領域の分割法は並列にアクセスしない場合のシーケンシャルにアクセスする順番を並列度に合わせて等分に分割するものとする。

【０１８２】この並列アクセスによる分割後の１つのアクセス領域を全て同一の物理記憶装置１８上に配置するのは必ずしも現実的ではない。そこで、分割後のアクセス領域がある一定量以上連続にまとまって１つの物理記憶装置上に記憶されていればよいと判断する。ただし、どのような場合でも連続してアクセスされることがなく、分割後のアクセス領域が異なるものに分類されるものに関しては、並列シーケンシャルアクセス時にアクセスがぶつかる可能性があるため、異なる物理記憶装置１８に記憶するという指針を設けて、これに沿うようなデータ配置を作成することによりシーケンシャルアクセスの性能を高める。

【０１８３】ステップ２６０１で処理を開始する。ステップ２６０２において、指定された表の物理記憶装置上に割り当てられた記憶領域利用総量を求める。この値は、ＤＢＭＳデータ構造情報６２１中のデータ構造データ量６４１を参照することにより求める。ステップ２６０３においては、ＤＢＭＳデータ構造情報６２１を参照して指定データ構造における最大アクセス並列度５６９を取得する。

【０１８４】ステップ２６０４において、ステップ２６０２で求めた指定表の記憶領域利用総量をステップ２６０３で求めた最大アクセス並列度５６９で割った量が、並列アクセス時にシーケンシャルにアクセスされる１つの領域のデータ量である。データ構造物理記憶位置情報７１２はシーケンシャルアクセス実行順にソートされているため、これを用いて最大アクセス並列度５６９の並列アクセスが実行されると仮定した前述のデータ分割指針を作成する。

【０１８５】ステップ２６０５において、データ再配置ワーク情報６７０を参照しながら、指定データ構造はステップ２６０４で作成したデータ分割指針に沿ったデータ配置が物理記憶装置１８上で行われているか確認し、そうであればステップ２６０９に進み、データ再配置案作成処理成功として処理を終了する。そうでない場合にはステップ２６０６に進む。

【０１８６】ステップ２６０６においては、物理記憶装置１８上において、ステップ２６０４で求めたデータ分割指針に従うデータ配置を求める。このとき、データがある一定値以下の領域に細分化されている場合には、連続した空き領域を探し、そこにアクセス構造を保つようにデータを移動するデータ移動案を作成する。また、最大アクセス並列度５６９の並列アクセスにより異なるアクセス領域に分離されるデータが同じ物理記憶装置１８上に配置されないようなデータ移動案を作成する。

【０１８７】ステップ２６０７においては、ステップ２６０６のデータ移動案作成処理が成功したか確認をする。成功した場合にはステップ２６０８に進み、失敗した場合にはステップ２６１０に進み、データ再配置案作成処理失敗として処理を終了する。

【０１８８】ステップ２６０８においては作成したデータ移動案を移動プラン情報７５０に記憶し、ステップ２６０９に進みデータ再配置案作成処理成功として処理を終了する。＜第二の実施の形態＞本実施の形態では、ＤＢＭＳが実行される計算機とファイルを管理単位とする記憶装置がネットワークを用いて接続された計算機システムにおいて、記憶装置がＤＢＭＳに関する情報、記憶装置外におけるデータの記憶位置のマッピングに関する情報を取得し、それらを用いて記憶装置の動作を改善する。

【０１８９】記憶装置において、記憶装置内部でデータの記憶位置を動的に変更する機能を有し、取得した情報をもとに好適なデータ再配置案を作成し、データの記憶位置の動的変更機能を用いて、作成したデータ再配置案に従ったデータ配置を実現し、アクセス性能を改善する。また、取得情報をもとにしたデータキャッシュの制御を行いより良いアクセス性能特性が得られるようにする。

【０１９０】図２６は、本発明の第二の実施の形態における計算機システムの構成図である。図示されたように、本発明の第二の実施の形態は本発明の第一の実施の形態と以下の点が異なる。

【０１９１】本実施の形態においてはＩ／Ｏパスインターフェイス７０、Ｉ／Ｏパス７１、Ｉ／Ｏパススイッチ７２が存在せず、記憶制御装置１０ｂとＤＢホスト８０ｃ，８０ｄはネットワーク７９を介してのみ接続される。記憶装置１０はファイルを単位としたデータ記憶管理を行う記憶装置１０ｂに変更される。そのため、物理記憶装置稼動情報３２、データキャッシュ管理情報３４、ＤＢＭＳデータ情報３６、ボリューム物理記憶位置管理情報３８がそれぞれ物理記憶装置稼動情報３２ｂ、データキャッシュ管理情報３４ｂ、ＤＢＭＳデータ情報３６ｂ、ファイル記憶管理情報３８ｂに変更される。

【０１９２】ＤＢホスト８０ｃ，８０ｄで実行されるＯＳ１００ではボリュームマネージャ１０２、ファイルシステム１０４が削除されその代わりに記憶装置１０ｂが提供するファイルをアクセスするための機能を有するネットワークファイルシステム１０４ｂが追加され、ＯＳ１００が保持するマッピング情報１０６がネットワークマウント情報１０６ｂへ変更される。

【０１９３】記憶装置１０はファイルを管理単位とする記憶装置１０ｂに変更される。ＤＢホスト８０ｃ，８０ｄからのアクセスもＮＦＳ等のファイルをベースとしたプロトコルで実施される。記憶装置１０におけるボリュームの役割は、記憶装置１０ｂにおいてはファイルもしくはファイルを管理するファイルシステムとなり、そのファイルの記憶位置管理情報がファイル記憶管理情報３

８ｂである。１つの記憶装置１０ｂの中に複数のファイルシステムが存在しても構わない。物理記憶装置１８の稼動情報はボリュームを単位とした取得からファイルシステムまたはファイルを単位とした取得に変更する。記憶装置１０ｂ内にファイルシステムが存在する場合でもデータの移動機能を実現可能である。

【０１９４】図２７はＤＢホスト８０ｃ，８０ｄのＯＳ１００内に記憶されているネットワークマウント情報１０６ｂを示す。ネットワークマウント情報１０６ｂは、記憶装置１０ｂから提供され、ＤＢホスト８０ｃ，８０ｄにおいてマウントされているファイルシステムの情報で、ファイルシステムの提供元記憶装置とそのファイルシステムの識別子である記憶装置名５８３とファイルシステム名１００１、そして、そのファイルシステムのマウントポイントの情報であるマウントポイント１０３１の組を保持する。

【０１９５】図２８は記憶装置１０ｂ内に保持されるファイル記憶管理情報３８ｂを示す。図５のボリューム物理記憶位置管理情報３８からの変更点は、ボリューム物理記憶位置メイン情報５１０、ボリュームデータ移動管理情報５１１からファイル物理記憶位置情報５１０ｂ、ファイルデータ移動管理情報５１１ｂにそれぞれ変更される。上記の変更内容は、ボリュームの識別子であるボリューム名５０１がファイルの識別子となるファイルシステム名１００１とファイルパス名１００２に、ボリューム内のデータ領域を示すボリューム論理ブロック番号５１２と移動論理ブロック番号７８２がそれぞれファイルブロック番号１００３または移動ファイルブロック番号１０２１に変更されるものである。

【０１９６】ここで、ファイルパス名１００２が“Ｅｍｐｔｙ”であるエントリ１０１５は特殊なエントリであり、このエントリには記憶装置１０ｂ内の物理記憶装置１８の領域のうち、指定ファイルシステム内でファイルの記憶領域として割り当てられていない領域を示し、図５中のボリュームデータ移動管理情報５１１を用いるデータ移動方式で説明した処理手順を用い、この領域に対して移動するデータをコピーすることによりデータの物理記憶位置の動的変更機能を実現する。

【０１９７】ここで注意が必要なのは、データ移動案作成時にデータ移動先の制約が増えた点である。本実施の形態においては、ファイルシステムを複数保持することが許されている。一般のファイルシステムにおいては、あるファイルシステムが他のファイルシステムが管理する領域を利用することは不可能である。つまり、一般のファイルシステムを用いている場合には、ファイルの移動は、そのファイルが存在しているファイルシステム内に閉じる必要がある。ただし、あるファイルシステムが他のファイルシステムが管理する領域を利用可能な機構を有している場合にはこの限りではない。

【０１９８】図２９に記憶装置１０ｂ内に保持される物理記憶装置稼動情報３２ｂを示す。図６の物理記憶装置稼動情報３２からの変更点は、稼動情報取得単位がボリュームからファイルシステムに変更されたため、ボリューム名５０１の部分がファイルシステム名１００１に変更されたことである。また、稼動情報取得単位をファイルとしてもよく、このときはボリューム名５０１の部分がファイルシステム名１００１とファイルパス名１００２に変更される。

【０１９９】図３０に記憶装置１０ｂ内に保持されているＤＢＭＳデータ情報３６ｂを示す。図７のＤＢＭＳデータ情報３６からの変更点は、ボリュームを利用した記憶管理からファイル利用した記憶管理に変更されたためデータ構造物理記憶位置情報７１２に修正が加えられ、データ構造物理記憶位置情報７１２ｂに変更されたことである。

【０２００】図３１にＤＢＭＳデータ情報３６ｂ中に含まれるデータ構造物理記憶位置情報７１２ｂを示す。図９のデータ構造物理記憶位置情報７１２からの変更点は、ボリュームを利用した記憶管理からファイル利用した記憶管理に変更されたため、ボリューム名５０１とボリューム論理ブロック番号５１２の部分がファイルシステム名１００１とファイルパス名１００２とファイルブロック番号１００３に変更されたことである。この情報は、ＤＢＭＳデータ記憶位置情報６２２とネットワークマウント情報１０６ｂを記憶装置１０の外部から取得し、さらにファイル物理記憶位置情報５１０ｂを参照して、対応する部分を組み合わせることにより作成する。

【０２０１】図３２に記憶装置１０ｂ内に保持されているデータキャッシュ管理情報３４ｂを示す。図１３のデータキャッシュ管理情報３４からの変更点は、ボリュームを利用した記憶管理からファイル利用した記憶管理に変更されたため、キャッシュセグメント情報７２０に修正が加えられ、キャッシュセグメント情報７２０ｂに変更されたことである。キャッシュセグメント情報７２０ｂのキャッシュセグメント情報７２０からの変更点は、上述の理由により、ボリューム名５０１とボリューム論理ブロック番号５１２の部分がファイルシステム名１００１とファイルパス名１００２とファイルブロック番号１００３に変更されたことである。

【０２０２】図３３にステップ２００３において作成する情報であるデータ再配置ワーク情報６７０ｂを示す。図１８のデータ再配置ワーク情報６７０からの変更点は、ボリュームを利用した記憶管理からファイル利用した記憶管理に変更されたため、空き領域情報６８０とデータ構造物理記憶位置情報７１２に修正が加えられ、それぞれ空き領域情報６８０ｂとデータ構造物理記憶位置情報７１２ｂへ変更されたことである。空き領域情報６８０ｂの空き領域情報６８０からの変更点は、ファイルシステムを利用した領域管理を実施しているため、空き領域管理はファイルシステムを意識する必要があり、空

き領域情報としては、データの記憶に利用していない場所を示す物理記憶装置名５０２と物理ブロック番号５１４とその空き領域を管理するファイルシステム名１００１の組を保持する。空き領域情報６８０ｂはファイル物理記憶位置メイン情報５１０ｂ中のファイルパス名１００１が“Ｅｍｐｔｙ”である領域を集めることにより初期化する。

【０２０３】図３４はステップ２００４で実行されるデータ配置解析・データ再配置案作成処理により作成されるデータ移動案を格納する移動プラン情報７５０ｂを示す。図１９の移動プラン情報７５０からの変更点は、ボリュームを利用した記憶管理からファイル利用した記憶管理に変更されたため、移動ボリューム名５６８と移動ボリューム論理ブロック番号７６９の部分が移動ファイルシステム名１１０１と移動ファイルパス名１１０２と移動ファイルブロック番号１１０３に変更されたことである。

【０２０４】前述のように、記憶装置１０において一般のファイルシステムを用いている場合には、ファイルの移動は、そのファイルが存在しているファイルシステム内に閉じる必要がある。従って、データ再配置案作成処理に関して、本実施の形態における第一の実施の形態から変更点は、データ移動先は現在データが存在しているファイルシステム上に限られるという制約が追加される。ただし、この制約も記憶装置１０が利用しているファイルシステムが他のファイルシステムが管理する領域を利用可能な機構を有している場合には除かれる。

【０２０５】記憶装置１０ｂにおける本実施の形態における第一の実施の形態からの差は、ほとんどがボリューム名５０１をファイルシステム名１００１とファイルパス名１００２に、ボリューム論理ブロック番号５１２をファイルブロック番号１００３に変更することであり、その他の変更点もその差を述べてきた。記憶装置１０ｂにおける処理に関しては、前記のデータ再配置案作成処理における制約を除き、基本的にこれまで述べてきた変更点と同じ変更点への対応方法を実施すれば、第一の実施の形態における処理をほぼそのまま本実施の形態に当てはめることができる。

【０２０６】

【発明の効果】本発明により以下のことが可能となる。第一に、ＤＢＭＳが管理するデータを保持する記憶装置において、ＤＢＭＳの処理の特徴を考慮することによりＤＢＭＳに対してより好ましい性能特性を持つことができる。この記憶装置を用いることにより、既存のＤＢＭＳに対してプログラムの修正無しにＤＢＭＳ稼動システムの性能を向上させることができるようになる。つまり、高性能なＤＢシステムを容易に構築できるようになる。

【０２０７】第二に、記憶装置の性能最適化機能を提供するため、それにより記憶装置の性能に関する管理コストを削減することができる。特に、本発明は、ＤＢシステムの高性能化に寄与するため、この記憶装置を用いたＤＢシステムの性能に関する管理コストを削減することができる。更に、本発明を用いた記憶装置は、自動でＤＢＭＳの特性を考慮したデータ配置の改善を行うことができ、管理コストの削減に大きく寄与する。

【図面の簡単な説明】

【図１】第一の実施の形態における計算機システムの構成を示す図である。

【図２】ＤＢホスト８０ａ，８０ｂのＯＳ１００内に記憶されているマッピング情報１０６を示す図である。

【図３】ＤＢＭＳ１１０ａ，１１０ｂ内に記憶されているその内部で定義・管理しているデータその他の管理情報であるスキーマ情報１１４を示す図である。

【図４】ＤＢホスト８０ａ，８０ｂのメモリ８８上に記憶されている実行履歴情報１２２を示す図である。

【図５】記憶装置１０内に保持されているボリューム物理記憶位置管理情報３８を示す図である。

【図６】記憶装置１０内に保持されている物理記憶装置稼動情報３２を示す図である。

【図７】記憶装置１０内に保持されているＤＢＭＳデータ情報３６を示す図である。

【図８】ＤＢＭＳデータ情報３６中に含まれるＤＢＭＳスキーマ情報７１１を示す図である。

【図９】ＤＢＭＳデータ情報３６中に含まれるデータ構造物理記憶位置情報７１２を示す図である。

【図１０】ＤＢＭＳデータ情報３６中に含まれるクエリ実行同時アクセスデータ構造カウント情報７１４を示す図である。

【図１１】ＤＢＭＳデータ情報３６に含まれるＤＢＭＳデータ構造キャッシュ効果情報７１５を示す図である。

【図１２】記憶装置１０において指定されたデータ構造のデータをデータキャッシュに保持する効果があるかどうかの判断する処理のフローを示す図である。

【図１３】記憶装置１０内に保持されているデータキャッシュ管理情報３４を示す図である。

【図１４】記憶装置１０がホストからデータの読出し要求を受け取ったときの処理フローを示す図である。

【図１５】記憶装置１０がホストからデータの書き込み要求を受け取ったときの処理フローを示す図である。

【図１６】アクセス先のデータの内容に従いアクセス要求のあったデータを保持するセグメントを適当な管理リストに繋ぐ処理のフローを示す図である。

【図１７】記憶装置１０内で実施されるデータ再配置処理の処理フローを示す図である。

【図１８】データ配置解析・再配置案作成処理で利用するデータ再配置ワーク情報６７０を示す図である。

【図１９】データ配置解析・再配置案作成処理で作成されるデータ移動案を格納する移動プラン情報７５０を示す図である。

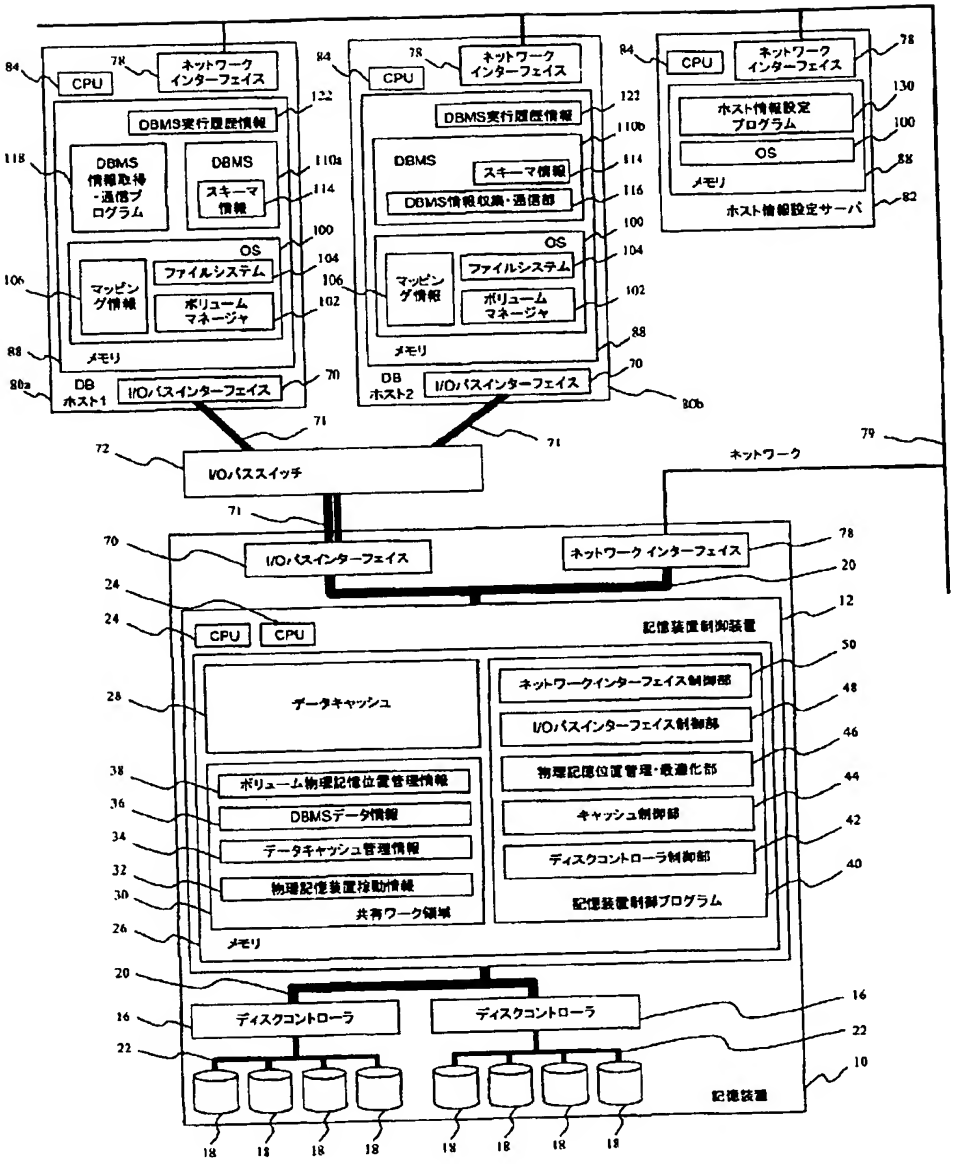
【図２０】物理記憶装置稼動情報３２を基にした同時アクセス実行データ構造を分離するためのデータ再配置案作成処理の処理フローを示す図である。
【図２１】クエリ実行時同時アクセスデータカウント情報７１４を利用する同時アクセス実行データ構造を分離するためのデータ再配置案作成処理処理フローを示す図である。
【図２２】指定されたデータ構造とそのデータ構造と同時にアクセスされる可能性が高いデータ構造の組を分離するデータ再配置案を作成する処理のフローを示す図である。
【図２３】データ構造の定義を基にした同時アクセス実行データ構造を分離するためのデータ再配置案作成処理の処理フローを示す図である。
【図２４】特定の表や索引の同一データ構造に対するアクセス並列度を考慮したデータ再配置案作成処理の処理フローを示す図である。
【図２５】特定の表データに対するシーケンシャルアクセス時のディスクネックを解消するデータ再配置案作成処理の処理フローを示す図である。
【図２６】第二の実施の形態における計算機システムの構成を示す図である。
【図２７】ＤＢホスト８０ｃ，８０ｄのＯＳ１００内に記憶されているネットワークマウント情報１０６ｂを示す図である。
【図２８】記憶装置１０ｂ内に保持されるファイル記憶管理情報３８ｂを示す図である。
【図２９】記憶装置１０ｂ内に保持される物理記憶装置稼動情報３２ｂを示す図である。
【図３０】記憶装置１０ｂ内に保持されているＤＢＭＳデータ情報３６ｂを示す図である。
【図３１】ＤＢＭＳデータ情報３６ｂ中に含まれるデータ構造物理記憶位置情報７１２ｂを示す図である。
【図３２】記憶装置１０ｂ内に保持されているデータキャッシュ管理情報３４ｂを示す図である。
【図３３】データ配置解析・再配置案作成処理で利用するデータ再配置ワーク情報６７０ｂを示す図である。
【図３４】データ配置解析・再配置案作成処理で作成されるデータ移動案を格納する移動プラン情報７５０ｂを示す図である。

【符号の説明】

１０，１０ｂ	記憶装置
１８	物理記憶装置
２８	データキャッシュ
３２，３２ｂ	物理記憶装置稼動情報
３４，３４ｂ	データキャッシュ管理情報
３６，３６ｂ	ＤＢＭＳデータ情報
３８	ボリューム物理記憶位置管理情報
３８ｂ	ファイル記憶管理情報
４０	記憶装置制御プログラム
４２	ディスクコントローラ制御部
４４	キャッシュ制御部
４６	物理記憶位置管理・最適化部
４８	Ｉ／Ｏパスインターフェイス制御部
５０	ネットワークインターフェイス制御部
７０	Ｉ／Ｏパスインターフェイス
７１	Ｉ／Ｏパス
７２	Ｉ／Ｏパススイッチ
７８	ネットワークインターフェイス
７９	ネットワーク
８０ａ，８０ｂ，８０ｃ，８０ｄ	ＤＢホスト
８２	ホスト情報設定サーバ
１００	ＯＳ（オペレーティングシステム）
１０２	ボリュームマネージャ
１０４	ファイルシステム
１０４ｂ	ネットワークファイルシステム
１０６	マッピング情報
１０６ｂ	ネットワークマウント情報
１１０ａ，１１０ｂ	ＤＢＭＳ（データベース管理システム）
１１４	スキーマ情報
１１６	ＤＢＭＳ情報通信部
１１８	ＤＢＭＳ情報取得・通信プログラム
１２２	実行履歴情報
１３０	ホスト情報設定プログラム

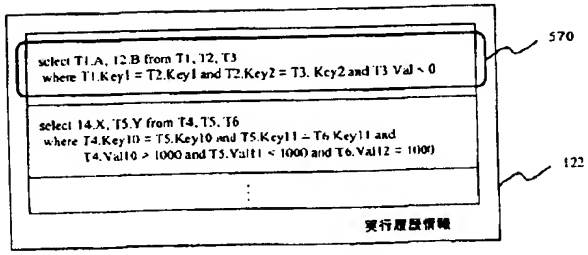
【図１】

図1



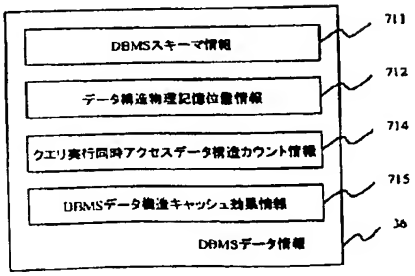
【図４】

図4



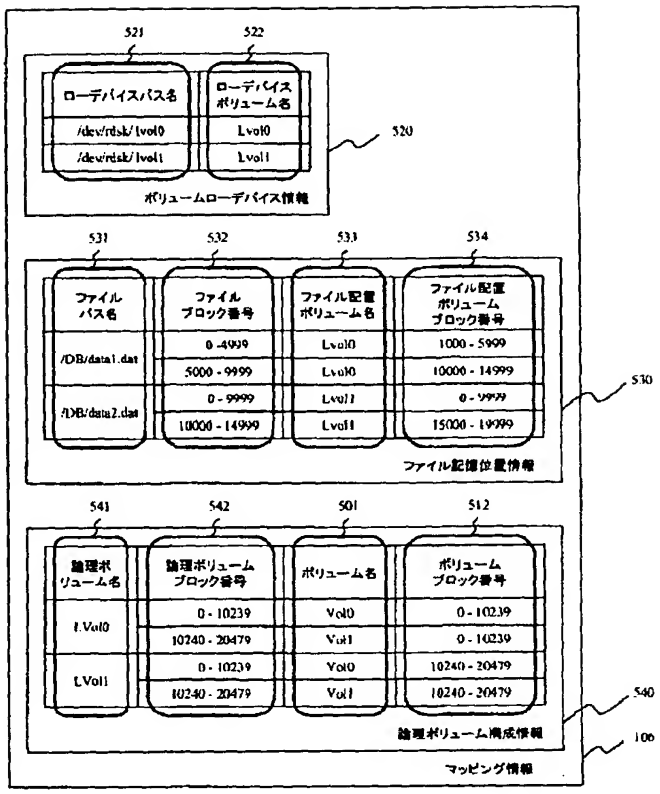
【図７】

図7



【図２】

図2



【図６】

図6

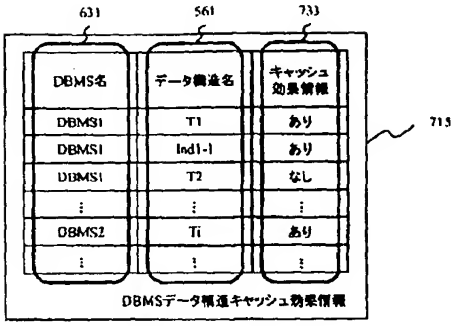
ボリューム名		Vol0	Vol0	Vol1	…	501
物理記憶装置名		Pdisk 0	Pdisk 1	Pdisk 0	…	502
累積稼動時間		23917390	38902849	8012891	…	503
旧累積稼動時間		22787638	38783484	7592039	…	593
稼働率	2000/4/1 12:00～2000/4/1 12:15	20%	12%	4%	…	594
	2000/4/1 12:15～2000/4/1 12:30	15%	10%	7%	…	
	2000/4/1 12:30～2000/4/1 12:45	16%	9%	5%	…	
	⋮	⋮	⋮	⋮	⋮	

595 前回累積稼動時間取得時刻：2001/4/12 18:15

32 物理記憶装置稼動情報

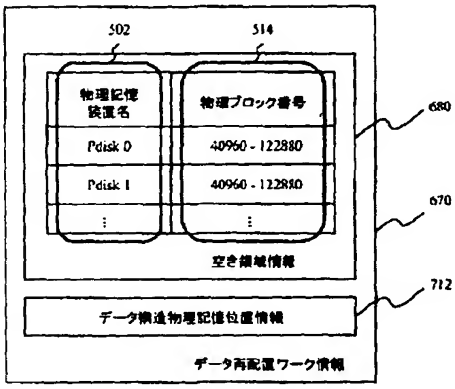
【図１１】

図11



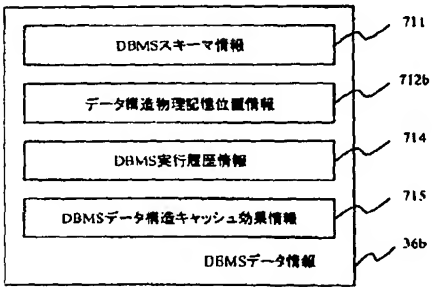
【図１８】

図18



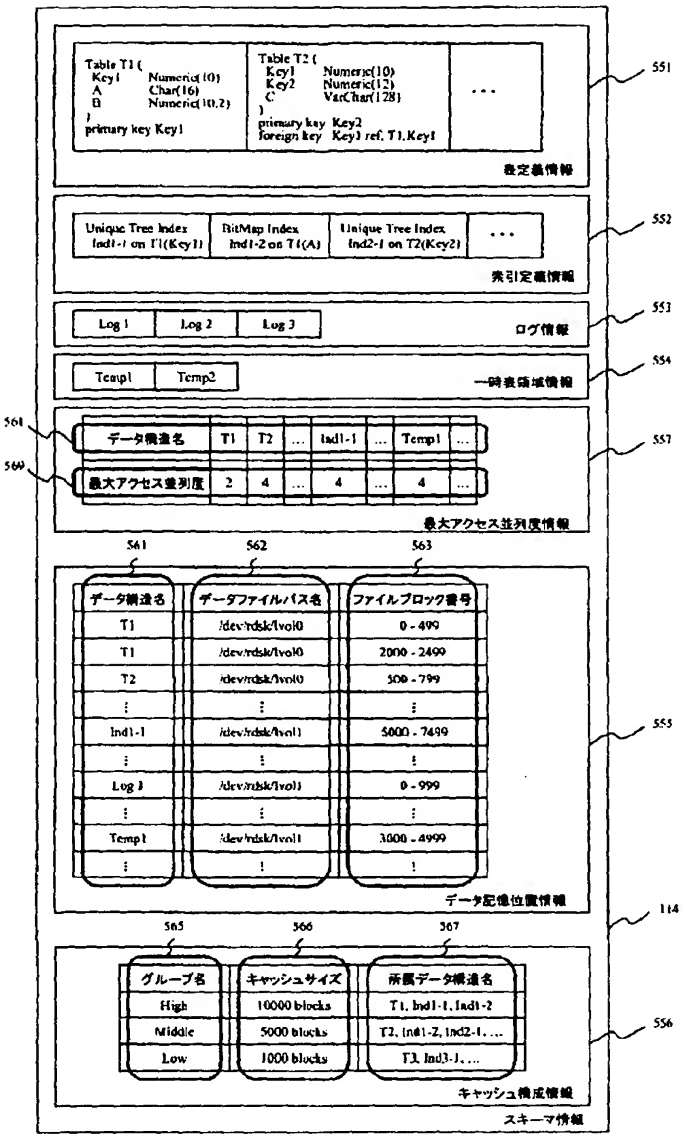
【図３０】

図30



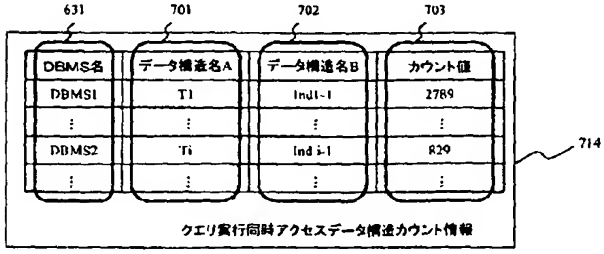
【図３】

図3



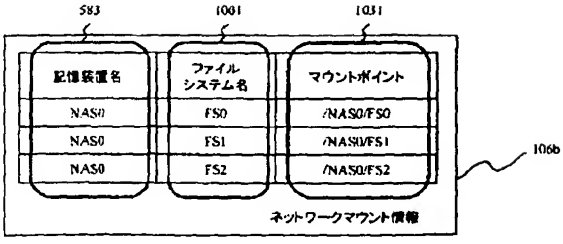
【図１０】

図10



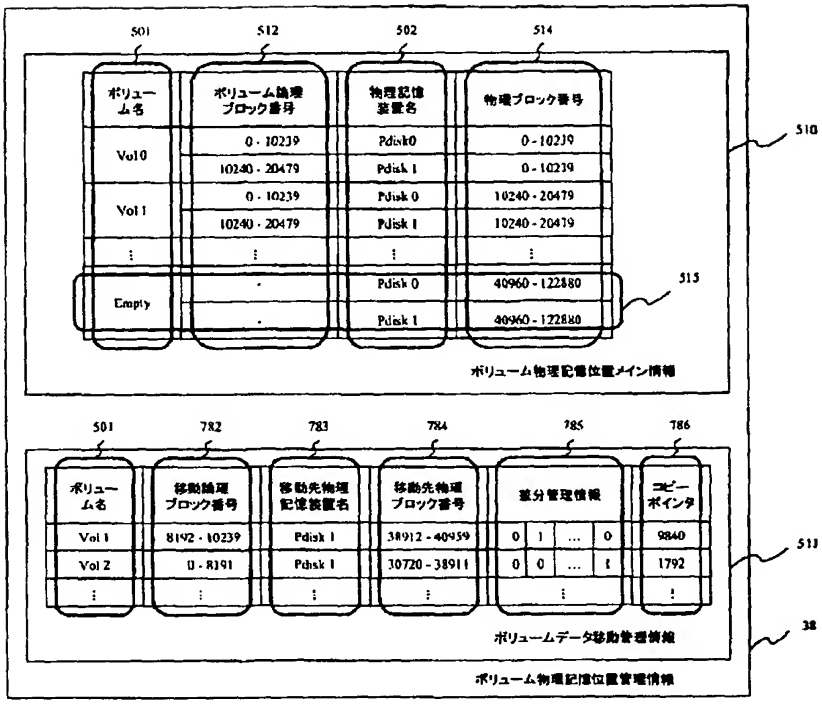
【図２７】

図27



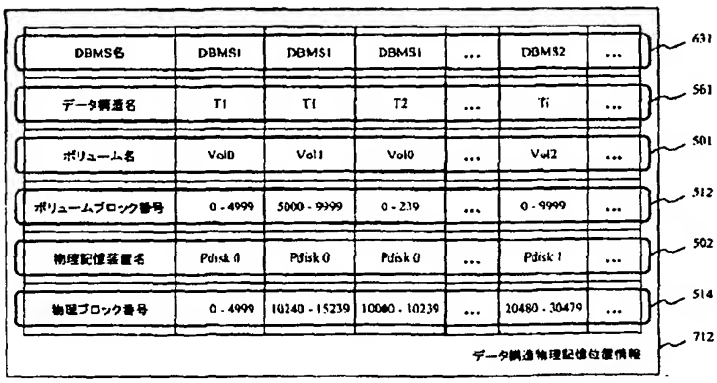
【図５】

図5


501
512
502
514
ボリューム名
ボリューム論理ブロック番号
物理記憶装置名
物理ブロック番号
Vol0
0 - 10239
Pdisk0
0 - 10239
10240 - 20479
Pdisk 1
0 - 10239
Vol 1
0 - 10239
Pdisk 0
10240 - 20479
10240 - 20479
Pdisk 1
10240 - 20479
Empty
Pdisk 0
40960 - 122880
Pdisk 1
40960 - 122880
510
515
ボリューム物理記憶位置メイン情報
501
782
783
784
785
786
ボリューム名
移動論理ブロック番号
移動先物理記憶装置名
移動先物理ブロック番号
差分管理情報
コピーポインタ
Vol 1
8192 - 10239
Pdisk 1
38912 - 40959
0 1 ... 0
9840
Vol 2
0 - 8191
Pdisk 1
30720 - 38911
0 0 ... 1
1792
511
ボリュームデータ移動管理情報
38
ボリューム物理記憶位置管理情報

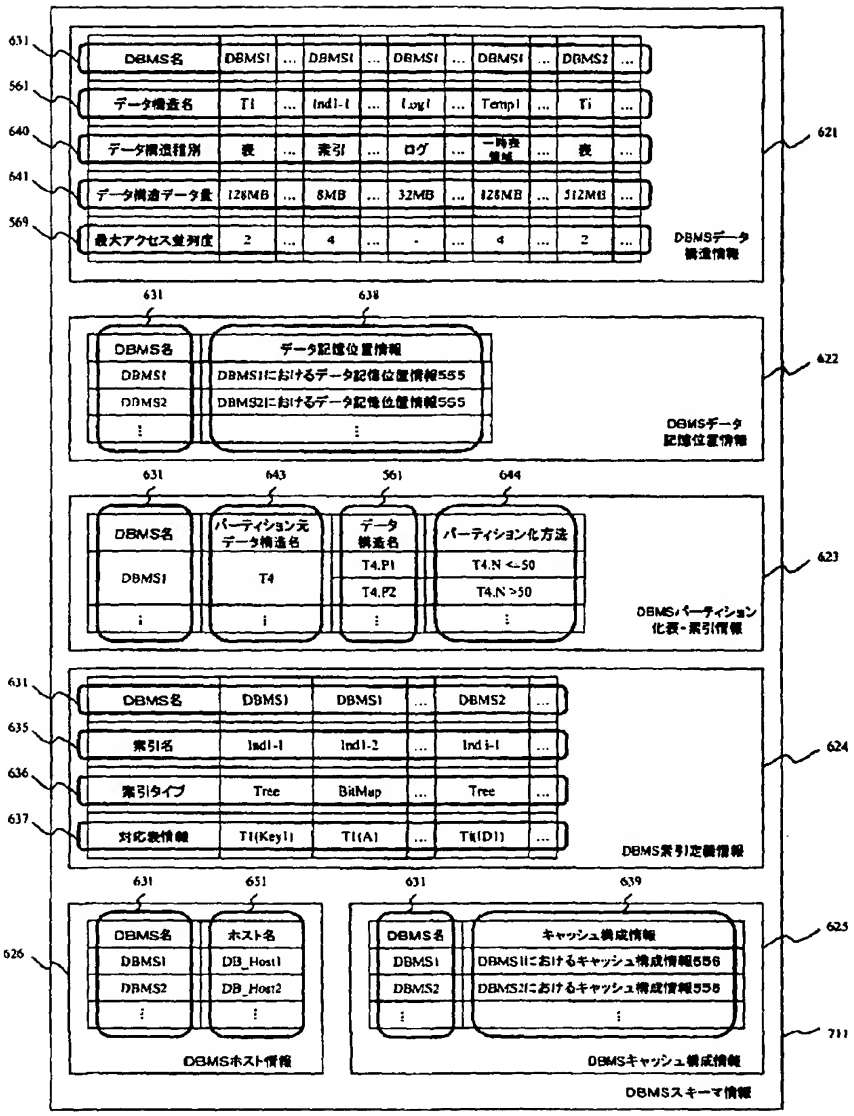
【図９】

図9


DBMS名
DBMS1
DBMS1
DBMS1
...
DBMS2
...
631
データ構造名
T1
T1
T2
...
Ti
...
561
ボリューム名
Vol0
Vol1
Vol0
...
Vol2
...
501
ボリュームブロック番号
0 - 4999
5000 - 9999
0 - 239
...
0 - 9999
...
512
物理記憶装置名
Pdisk 0
Pdisk 0
Pdisk 0
...
Pdisk 1
...
502
物理ブロック番号
0 - 4999
10240 - 15239
10080 - 10239
...
20480 - 30479
...
514
712
データ構造物理記憶位置情報

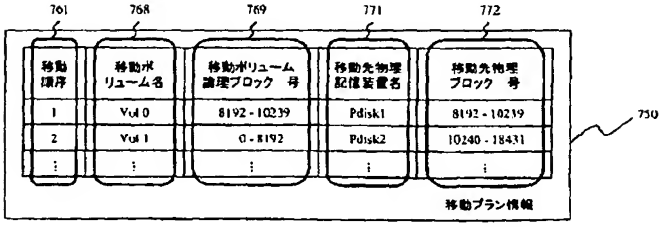
【図８】

図8


631
DBMS名
DBMS1
561
データ構造名
T1
640
データ構造種別
表
641
データ構造データ量
128MB
569
最大アクセス並列度
2
621
DBMSデータ構造情報
631
638
DBMS名
データ記憶位置情報
DBMS1
DBMS1におけるデータ記憶位置情報555
DBMS2
DBMS2におけるデータ記憶位置情報555
622
DBMSデータ記憶位置情報
631
643
561
644
DBMS名
パーティション元データ構造名
データ構造名
パーティション化方法
DBMS1
T4
T4.P1
T4.N <=50
T4.P2
T4.N >50
623
DBMSパーティション化表・索引情報
631
635
636
637
DBMS名
索引名
索引タイプ
対応表情報
DBMS1
Ind1-1
Ind1-2
Tree
BitMap
T1(Key1)
T1(A)
624
DBMS索引定義情報
631
651
631
639
DBMS名
ホスト名
DBMS1
DB_Host1
DBMS2
DB_Host2
626
DBMSホスト情報
DBMS名
キャッシュ構成情報
DBMS1
DBMS1におけるキャッシュ構成情報556
DBMS2
DBMS2におけるキャッシュ構成情報556
625
711
DBMSキャッシュ構成情報
DBMSスキーマ情報

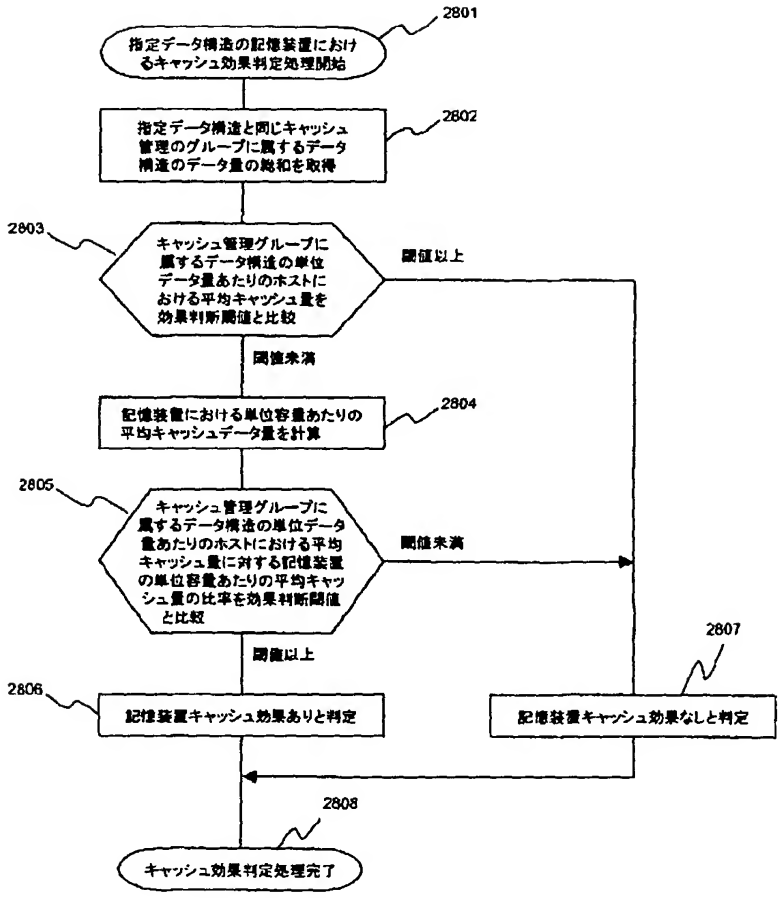
【図１９】

図19


761
768
769
771
772
移動順序
移動ボリューム名
移動ボリューム論理ブロック 号
移動先物理記憶装置名
移動先物理ブロック 号
1
Vol 0
8192 - 10239
Pdisk1
8192 - 10239
2
Vol 1
0 - 8192
Pdisk2
10240 - 18431
750
移動プラン情報

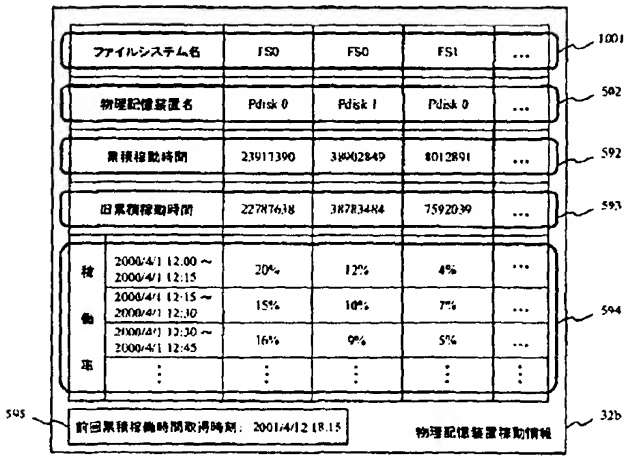
【図１２】

図12



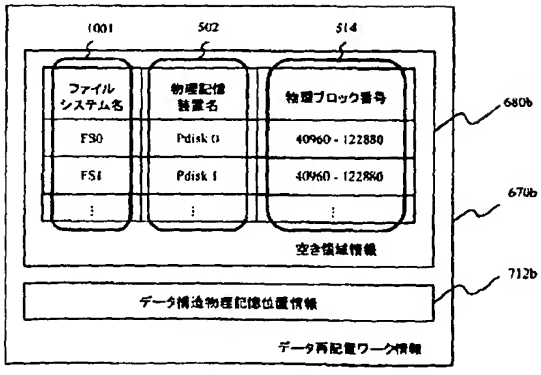
【図２９】

図29



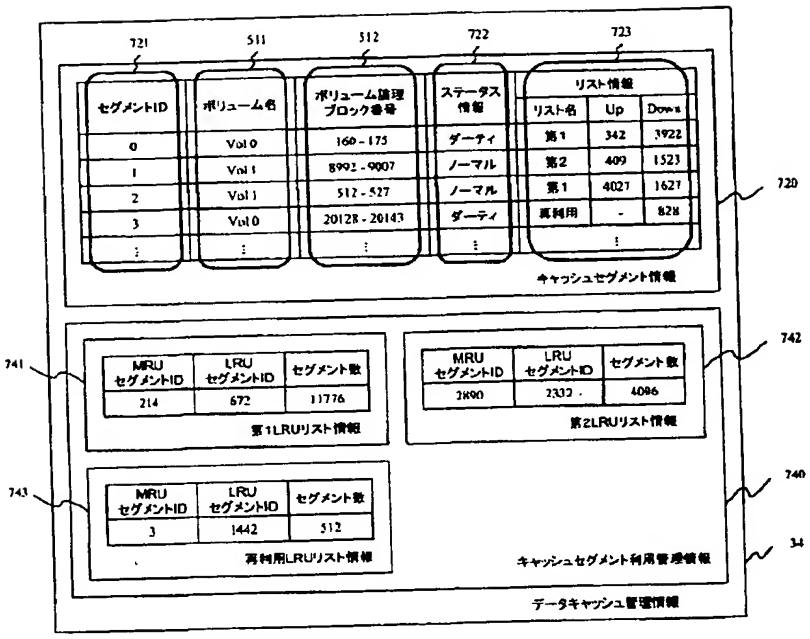
【図３３】

図33



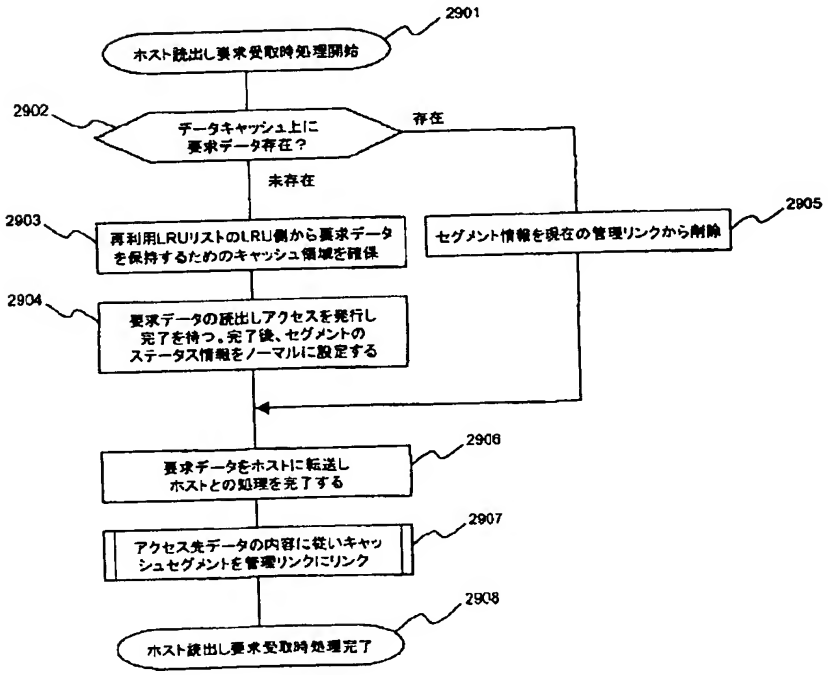
【図１３】

図13



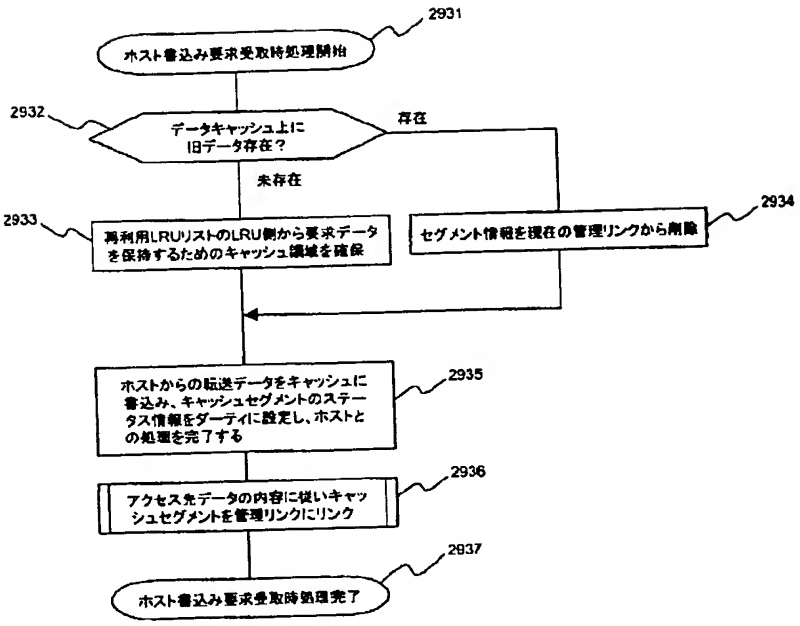
【図１４】

図14



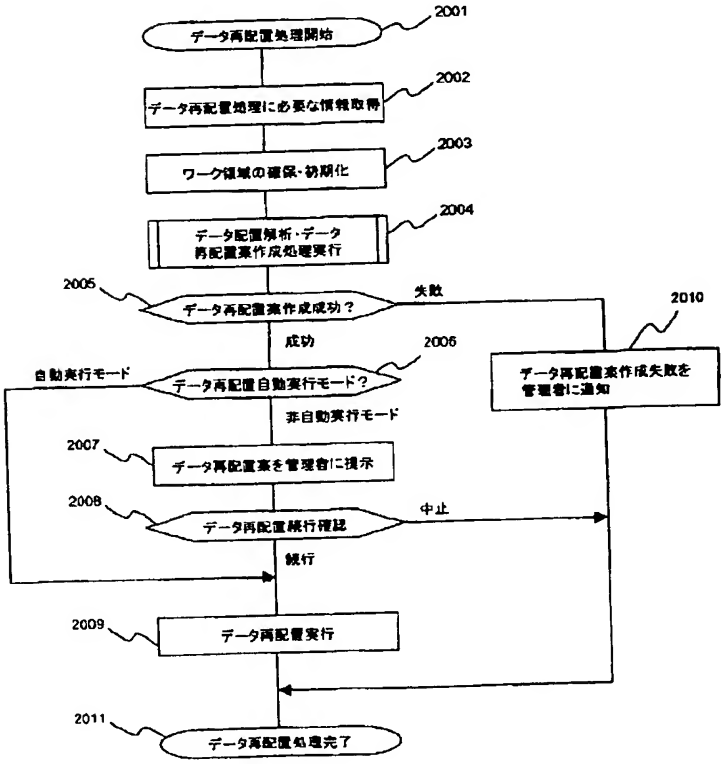
【図１５】

図15



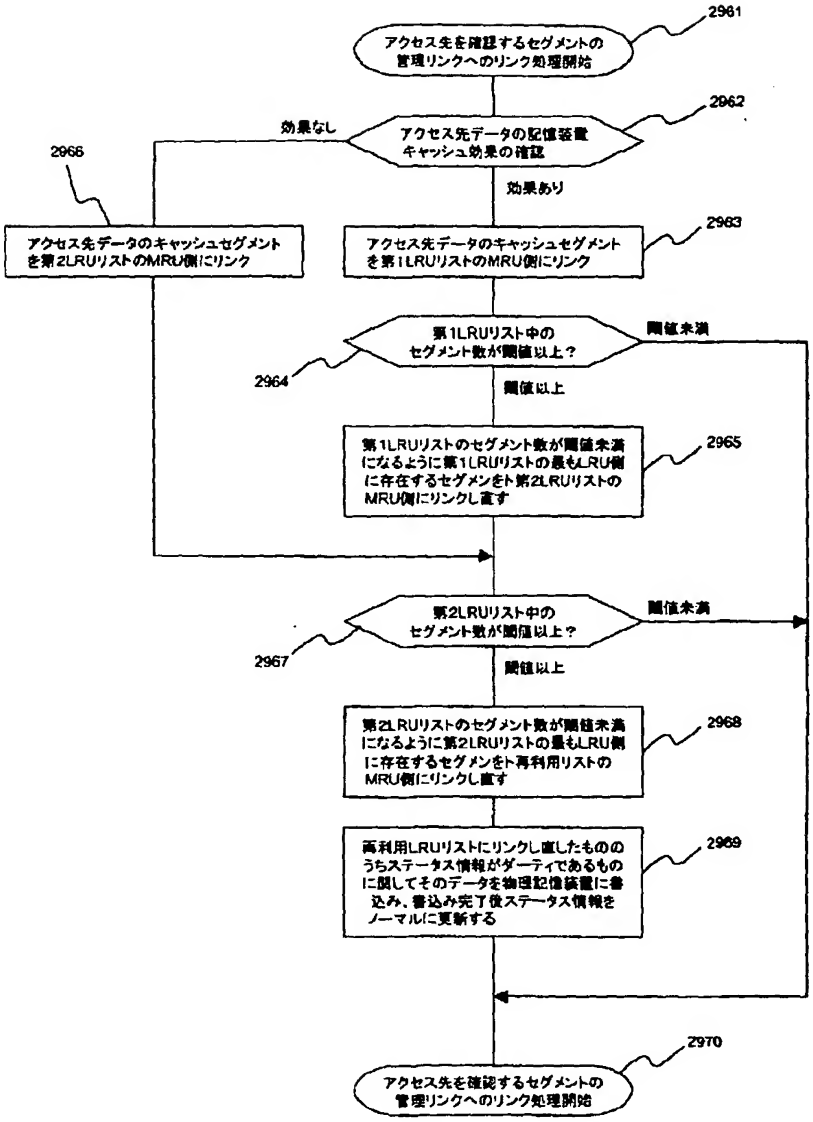
【図１７】

図17



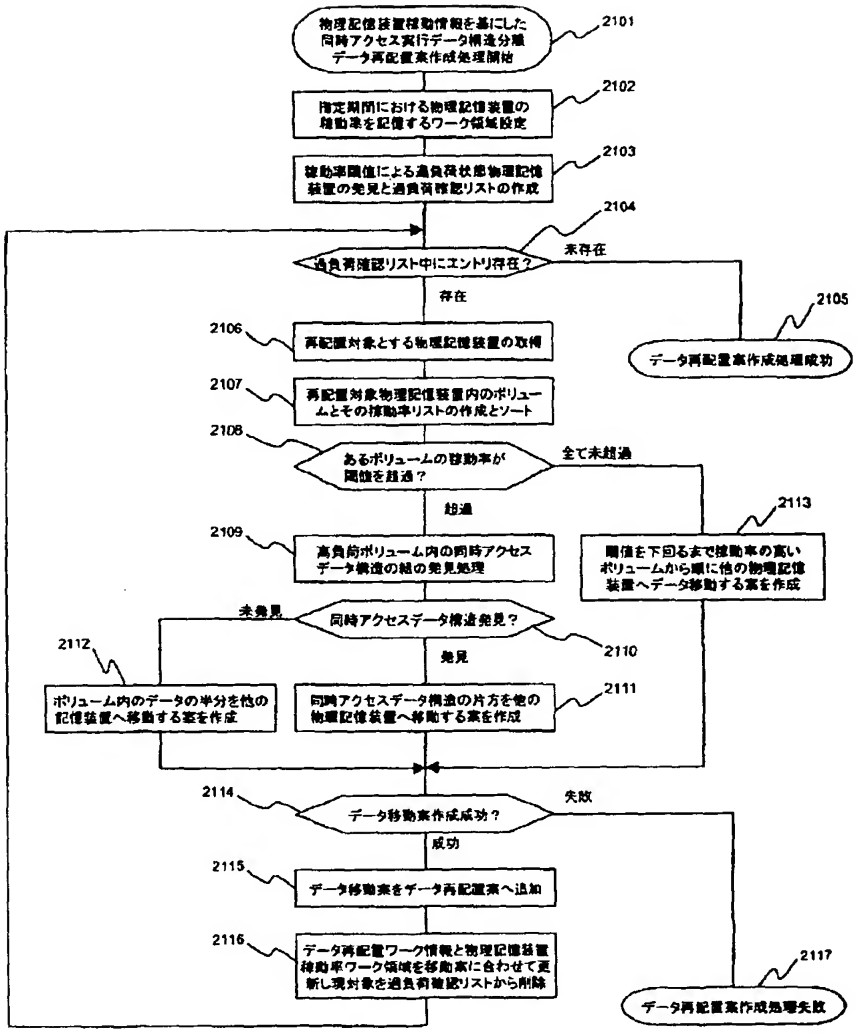
【図１６】

図16


2961
アクセス先を確認するセグメントの管理リンクへのリンク処理開始
2962
アクセス先データの記憶装置キャッシュ効果の確認
効果なし
効果あり
2966
アクセス先データのキャッシュセグメントを第2LRUリストのMRU側にリンク
2963
アクセス先データのキャッシュセグメントを第1LRUリストのMRU側にリンク
第1LRUリスト中のセグメント数が閾値以上？
閾値未満
2964
閾値以上
2965
第1LRUリストのセグメント数が閾値未満になるように第1LRUリストの最もLRU側に存在するセグメントを第2LRUリストのMRU側にリンクし直す
第2LRUリスト中のセグメント数が閾値以上？
閾値未満
2967
閾値以上
2968
第2LRUリストのセグメント数が閾値未満になるように第2LRUリストの最もLRU側に存在するセグメントを再利用リストのMRU側にリンクし直す
2969
再利用LRUリストにリンクし直したもののうちステータス情報がダーティであるものに関してそのデータを物理記憶装置に書込み、書込み完了後ステータス情報をノーマルに更新する
2970
アクセス先を確認するセグメントの管理リンクへのリンク処理開始

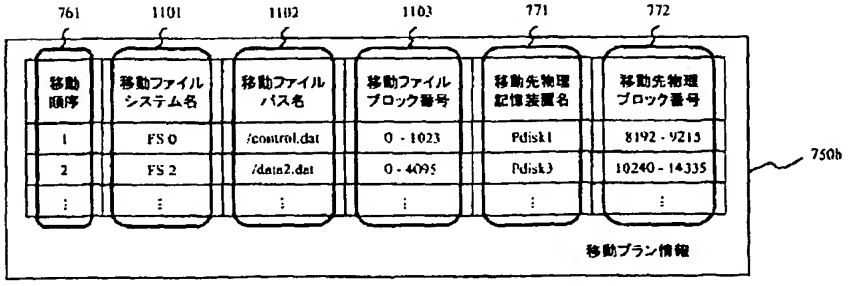
【図２０】

図20



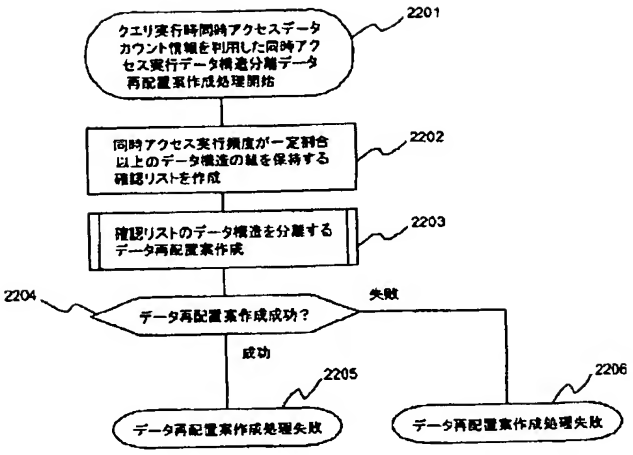
【図３４】

図34



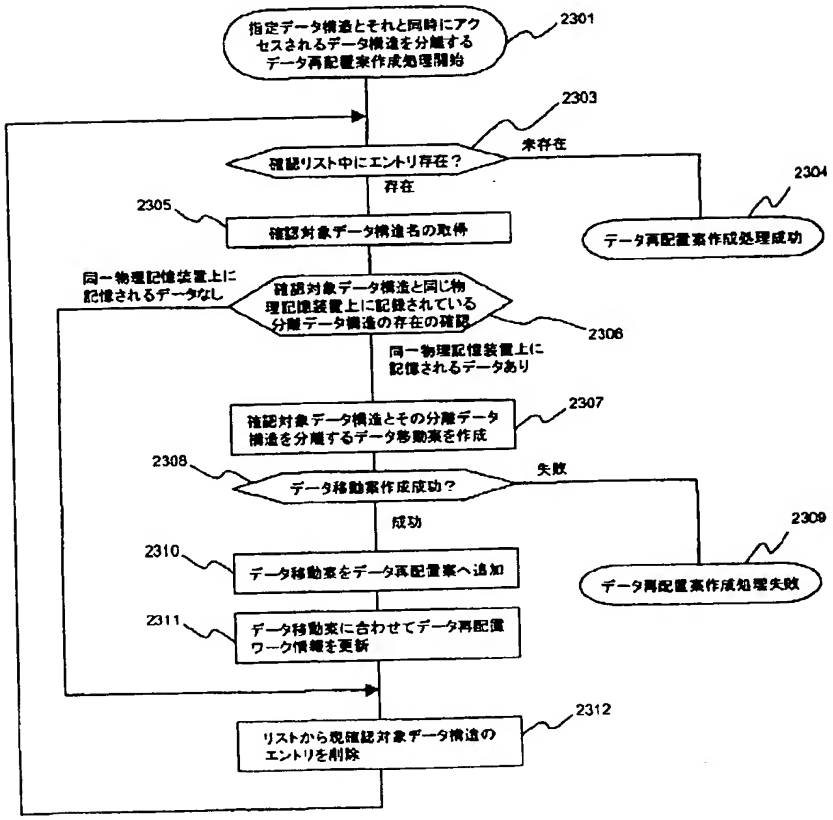
【図２１】

図21



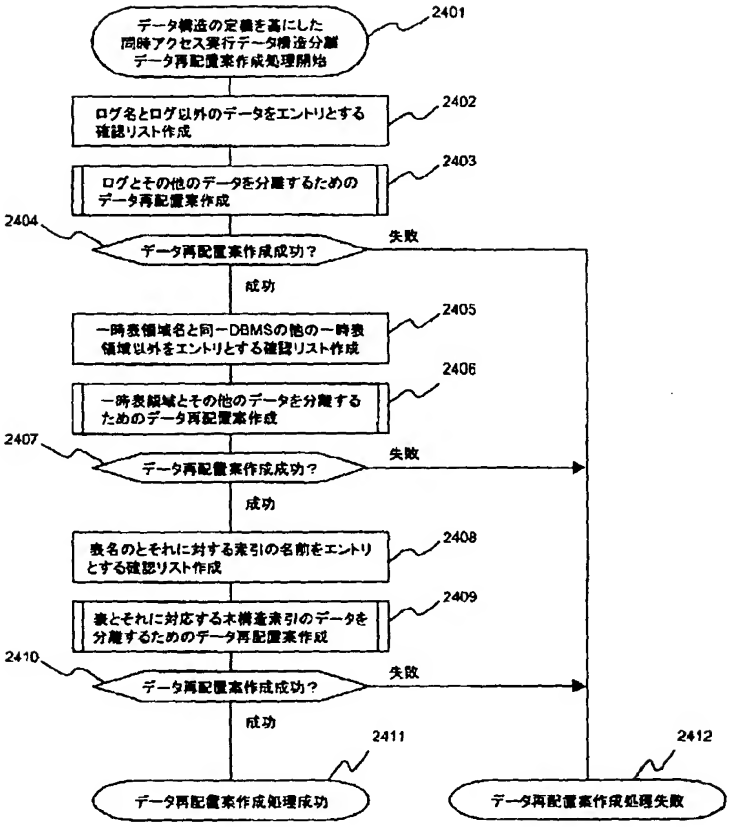
【図２２】

図22



【図２３】

図23



【図３１】

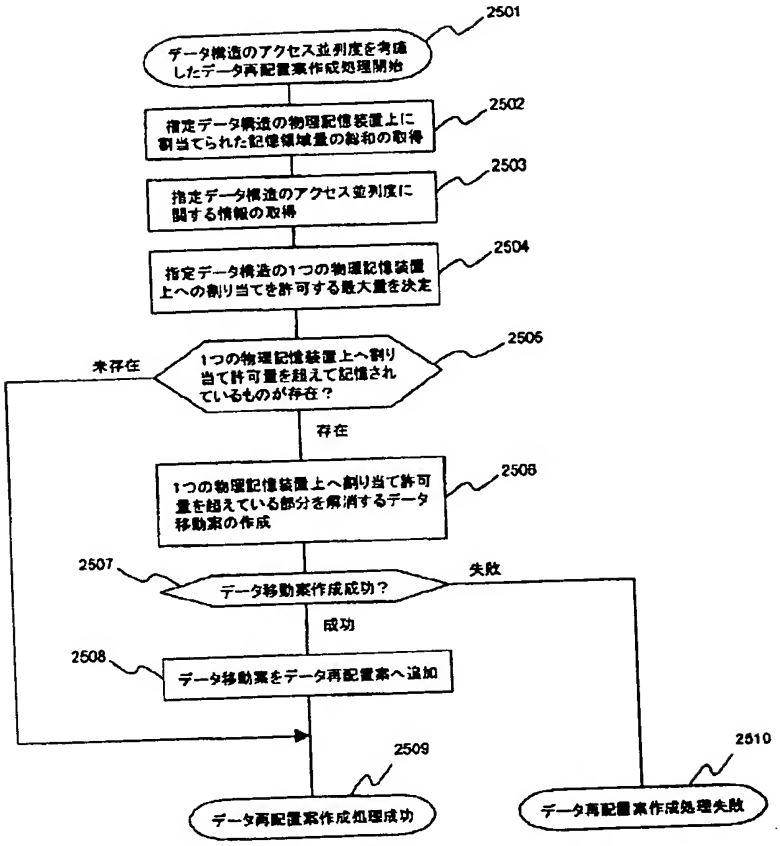
図31

DBMS名	DBMS1	DBMS1	DBMS1	…	DBMS2	…	631
データ構造名	T1	T1	T2	…	Ti	…	561
ファイルシステム名	FS1	FS1	FS1	…	FS2	…	1001
ファイルパス名	/data1.dat	/data1.dat	/data2.dat	…	/datai.dat	…	1002
ファイルブロック番号	0 - 4999	5000 - 9999	0 - 4999	…	0 - 9999	…	1003
物理記憶装置名	Pdisk 1	Pdisk 0	Pdisk 1	…	Pdisk 2	…	514
物理ブロック番号	10240 - 15239	10240 - 15239	15240-20239	…	20480 - 30479	…	502

データ構造物理記憶位置情報 7126

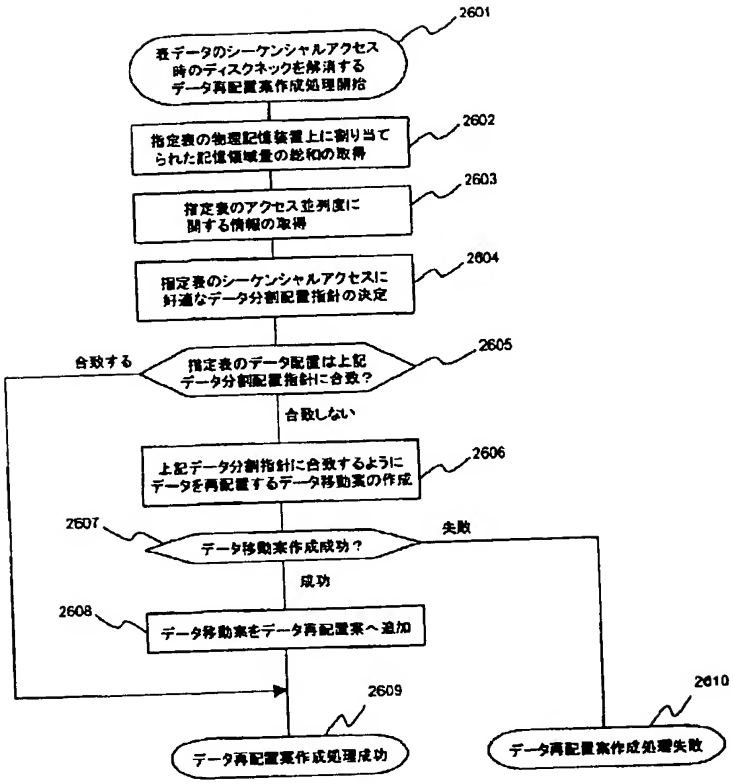
【図２４】

図24


2501
データ構造のアクセス並列度を考慮したデータ再配置案作成処理開始
2502
指定データ構造の物理記憶装置上に割当てられた記憶領域量の総和の取得
2503
指定データ構造のアクセス並列度に関する情報の取得
2504
指定データ構造の1つの物理記憶装置上への割り当てを許可する最大量を決定
2505
1つの物理記憶装置上へ割り当て許可量を超えて記憶されているものが存在？
未存在
存在
2506
1つの物理記憶装置上へ割り当て許可量を超えている部分を解消するデータ移動案の作成
2507
データ移動案作成成功？
失敗
成功
2508
データ移動案をデータ再配置案へ追加
2509
データ再配置案作成処理成功
2510
データ再配置案作成処理失敗

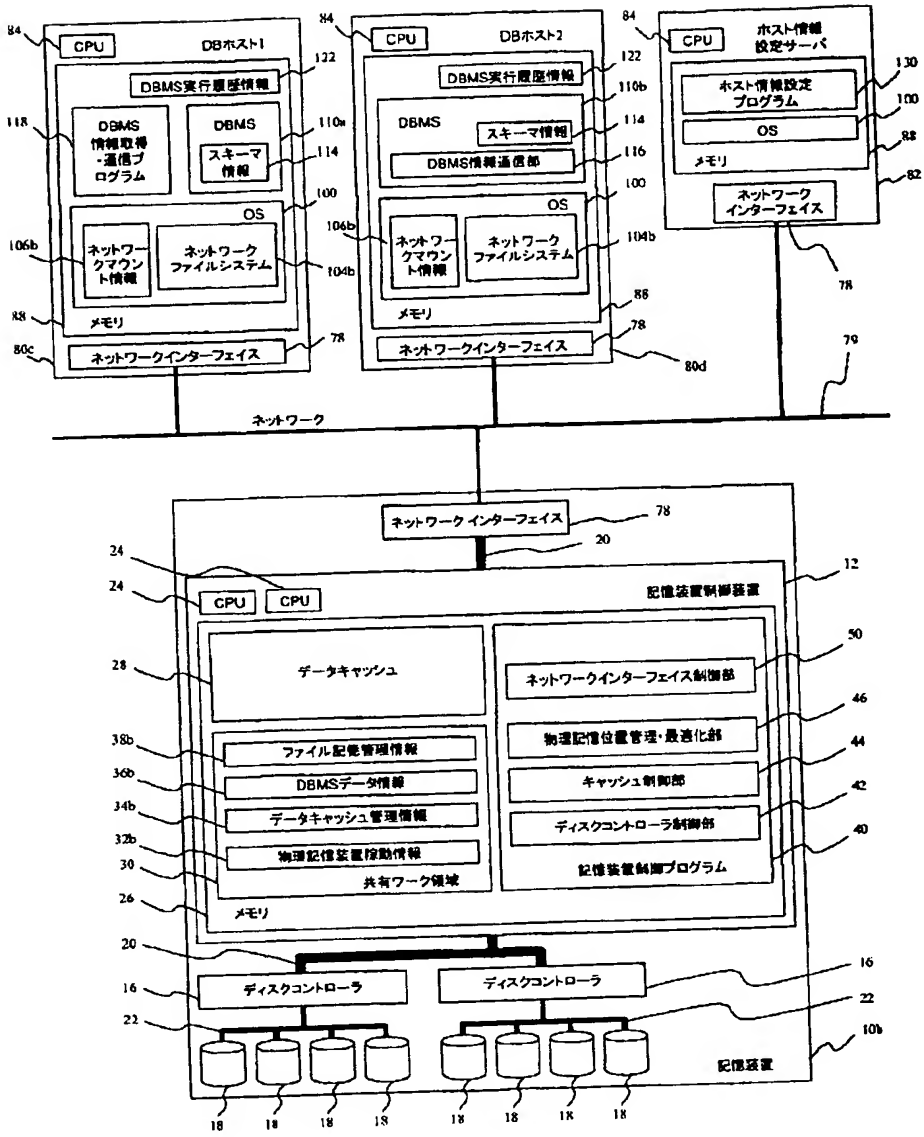
【図２５】

図25


2601
巨データのシーケンシャルアクセス時のディスクネックを解消するデータ再配置案作成処理開始
2602
指定表の物理記憶装置上に割り当てられた記憶領域量の総和の取得
2603
指定表のアクセス並列度に関する情報の取得
2604
指定表のシーケンシャルアクセスに好適なデータ分割配置指針の決定
2605
指定表のデータ配置は上記データ分割配置指針に合致？
合致する
合致しない
2606
上記データ分割指針に合致するようにデータを再配置するデータ移動案の作成
2607
データ移動案作成成功？
失敗
成功
2608
データ移動案をデータ再配置案へ追加
2609
データ再配置案作成処理成功
2610
データ再配置案作成処理失敗

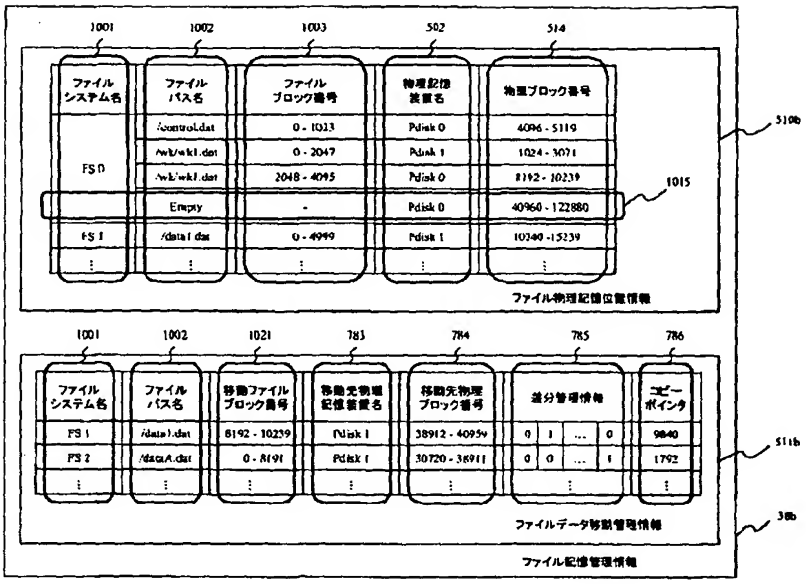
【図２６】

図26


CPU
DBホスト1
DBMS実行履歴情報
DBMS
情報取得
・通信プ
ログラム
DBMS
スキーマ
情報
OS
ネットワー
クマウン
ト情報
ネットワーク
ファイルシステム
メモリ
ネットワークインターフェイス
DBホスト2
スキーマ情報
DBMS情報通信部
ホスト情報
設定サーバ
ホスト情報設定
プログラム
ネットワーク
インターフェイス
ネットワーク
ネットワーク インターフェイス
記憶装置制御装置
データキャッシュ
ネットワークインターフェイス制御部
物理記憶位置管理・最適化部
キャッシュ制御部
ディスクコントローラ制御部
記憶装置制御プログラム
ファイル記憶管理情報
DBMSデータ情報
データキャッシュ管理情報
物理記憶装置稼動情報
共有ワーク領域
ディスクコントローラ
記憶装置

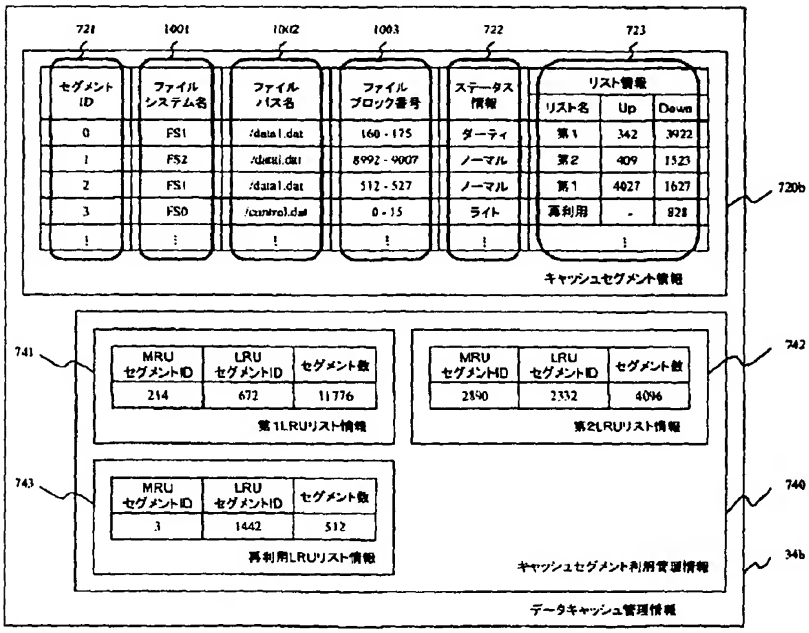
【図２８】

図28



【図３２】

図32



フロントページの続き

(72)発明者　喜連川　優
千葉県松戸市二十世紀が丘丸山町17

Ｆターム(参考)　5B005 JJ13 LL17 MM04 QQ02 QQ04
VV02
5B065 BA01 CA11 CC02 CC08 CE11
CH01 CH18
5B082 BA09 FA12